

OTTO-VON-GUERICKE-UNIVERSITÄT MAGDEBURG  
FAKULTÄT FÜR MATHEMATIK  
INSTITUT FÜR MATHEMATISCHE STOCHASTIK



# Stochastik für Ingenieure

– Eine Einführung in die Wahrscheinlichkeitstheorie und mathematische Statistik –  
(Vorlesungsskript Sommersemester 2004)

von

**Norbert Gaffke**

Online-Version: [www.math.uni-magdeburg.de/institute/imst/ag\\_gaffke](http://www.math.uni-magdeburg.de/institute/imst/ag_gaffke)



# Inhaltsverzeichnis

<b>1</b>	<b>Zufallsexperimente, W-Räume und Zufallsvariablen</b>	<b>3</b>
1.1	Zufallsexperiment und Modellierung . . . . .	3
1.2	W-Verteilung, W-Raum . . . . .	6
1.3	Zufallsvariablen . . . . .	7
1.4	Wichtige Klassen von diskreten Zufallsvariablen . . . . .	11
1.5	Gemeinsame Verteilung mehrerer Zufallsvariablen . . . . .	14
<b>2</b>	<b>Stoch. Unabhängigkeit, bedingte W'keiten</b>	<b>18</b>
2.1	Stochastische Unabhängigkeit von Ereignissen . . . . .	18
2.2	Stochastische Unabhängigkeit von Zufallsvariablen . . . . .	19
2.3	Modellierung unabhängiger Zufallsexperimente . . . . .	22
2.4	Bedingte Wahrscheinlichkeiten . . . . .	25
<b>3</b>	<b>Verteilungen reeller Zufallsvariablen</b>	<b>30</b>
3.1	Verteilungsfunktionen . . . . .	30
3.2	Diskrete Zufallsvariablen . . . . .	32
3.3	Stetig-verteilte Zufallsvariablen . . . . .	34
<b>4</b>	<b>Erwartungswert, Varianz und Standardabweichung</b>	<b>39</b>
4.1	Erwartungswert . . . . .	39
4.2	Standardabweichung und Varianz . . . . .	46
4.3	Varianzen von Linearkombinationen . . . . .	51
<b>5</b>	<b>Statistische Modelle und Fragestellungen</b>	<b>53</b>
5.1	Statistische Modellierung . . . . .	53
5.2	Spezielle Modelle . . . . .	55
5.3	Statistische Fragestellungen . . . . .	60

<b>6</b>	<b>Maximum-Likelihood-Schätzung</b>	<b>65</b>
6.1	Likelihood-Funktion . . . . .	65
6.2	Maximum-Likelihood-Schätzung . . . . .	68
<b>Anhang</b>		<b>77</b>
	Tafel A.1: Verteilungsfunktion einer standard-normalverteilten Zufallsvariablen . . . . .	77
	Tafel A.2: Quantile einer standard-normalverteilten Zufallsvariablen . . . . .	78
	Tafel A.3: Quantile einer t - verteilten Zufallsvariablen . . . . .	78
	Tafel A.4: Quantile einer $\chi^2$ - verteilten Zufallsvariablen . . . . .	79

# Kapitel 1

## Zufallsexperimente, W-Räume und Zufallsvariablen

### 1.1 Zufallsexperiment und Modellierung

Der Begriff *Zufallsexperiment* steht hier für jeden realen Vorgang, der vom Zufall beeinflusst wird. Typischerweise liefert ein Zufallsexperiment ein *Ergebnis*, das “zufällig” (zumindest teilweise) ist.

**Beispiele** für Zufallsexperimente:

- Glücksspiele (z.B. Münzwurf, Würfeln, Lotto)
- 0-1-Experimente (Bernoulli-Experimente), wobei z.B. “1” für *Erfolg* und “0” für *Misserfolg* steht (z.B. Therapie, Platzierung, Schießen). Das betrachtete Zufallsexperiment kann die ein-malige Durchführung eines 0-1-Experimentes sein oder auch eine mehr-malige (z.B. die 10-malige) unabhängige Durchführung eines 0-1-Experimentes.
- Zufällige Anzahlen (z.B. Anzahl von Kunden oder Jobs, Anzahl von Verkehrsunfällen, Anzahl radioaktiver Zerfälle)
- Lebensdauern (z.B. von technischen Bauteilen, von Lebewesen)
- Industrielle Fertigung: Charakteristika gefertigter Produkte (z.B. elektr. Widerstände, Festigkeit von Materialien, geometrische Abmessungen von techn. Bauteilen)

**Die mathematische Modellierung** eines Zufallsexperiments erfolgt durch

- (a) die Festlegung einer Menge  $M$ , die alle möglichen Ergebnisse des Zufallsexperiments enthält (das ist i.d.R. eine leichte Aufgabe);
- (b) die Festlegung einer passenden *Wahrscheinlichkeitsverteilung* (kurz: *W-Verteilung*) auf der Ergebnismenge  $M$  (das ist i.d.R. die schwierigere Aufgabe); das bedeutet vereinfacht gesagt:<sup>1</sup> Für jedes mögliche Ergebnis  $x \in M$  ist die Wahrscheinlichkeit seines Eintretens festzulegen.

---

<sup>1</sup>Streng genommen kann das nur im Fall einer diskreten Ergebnismenge  $M$ , z.B. einer endlichen Menge, so gesehen werden; für nicht-diskrete Ergebnismengen  $M$ , z.B.  $M = \mathbb{R}$ , ist die Situation komplizierter.

**Bemerkung 1.1**

Die Festlegung der Ergebnismenge  $M$  kann ruhig auch ‘Redundanzen’ beinhalten, d.h. Elemente  $x$  als ‘mögliche Ergebnisse’ einbeziehen, die praktisch gar nicht möglich sind (das hat manchmal gewisse Vorteile für die mathematischen Darstellungen). Die Gesamtheit solcher Elemente wird dann im Schritt (b) die Wahrscheinlichkeit Null erhalten, wodurch sie in der Tat als ein praktisch unmögliches Ereignis eingestuft ist. ■

Zur Ausführung von Schritt (b) verwenden wir den Begriff einer *Zufallsvariablen*. Wir fassen das Ergebnis  $x \in M$  des Zufallsexperiments als ein Wert einer Zufallsvariablen  $X$  auf. Mathematisch ist  $X$  eine Abbildung von einer gewissen Menge  $\Omega$  in die Menge  $M$ ,

$$X : \Omega \longrightarrow M ,$$

wobei die Menge  $\Omega$  nicht weiter spezifiziert werden braucht. Z.B. können wir uns die Menge  $\Omega$  als Menge aller möglichen Zustände einer ‘Welt’ vorstellen; der ‘Zufall’ greift einen Zustand  $\omega \in \Omega$  heraus; die Bedeutung der Abbildung  $X$  liegt darin, dass dann der Wert  $X(\omega)$  das Ergebnis  $x$  des betrachteten Zufallsexperiments ist. Aber auch der Zufall folgt einem gewissen System, was mathematisch so ausgedrückt wird: Es ist eine gewisse W-Verteilung  $P$  auf  $\Omega$  vorhanden (die präzise Definition dieses Begriffs geben wir in Abschnitt 1.2). Für die Modellierung des betrachteten Zufallsexperimentes ist die genaue Kenntnis dieser drei Objekte,  $\Omega$ ,  $P$  und  $X$  nicht erforderlich; relevant sind dafür nur die Wahrscheinlichkeiten der Form

$$P(X \in C) \quad (\text{für beliebige Teilmengen } C \subset M),$$

die Wahrscheinlichkeit, dass das Ergebnis des Zufallsexperiments in die Teilmenge  $C \subset M$  fällt. Diese Wahrscheinlichkeiten sind im Schritt (b) der Modellierung geeignet festzulegen. Im Fall einer diskreten Ergebnismenge  $M$  (z.B. einer endlichen Menge  $M$ ) genügt es, diese Festlegung für die einelementigen Teilmengen  $C$  zu treffen, d.h.  $C = \{x\}$  für  $x \in M$ ; statt “ $X \in \{x\}$ ” schreiben wir einfacher “ $X = x$ ”. Wir haben also geeignet festzulegen:

$$P(X = x) \quad (\text{für jedes einzelne } x \in M),$$

die Wahrscheinlichkeit, dass  $x$  das Ergebnis des Zufallsexperiments ist (für jedes einzelne  $x \in M$ ). Fassen wir zusammen:

Das Ergebnis eines Zufallsexperimentes wird als ein Wert  $x = X(\omega)$  einer Zufallsvariablen  $X : \Omega \longrightarrow M$  aufgefasst, wobei die Menge  $M$  alle möglichen Ergebnisse des Zufallsexperimentes beinhaltet. Wir gehen davon aus, dass auf dem Definitionsbereich  $\Omega$  der Zufallsvariablen eine W-Verteilung  $P$  vorliegt. Von entscheidender Bedeutung für die Modellbildung ist – neben der Festlegung der Menge  $M$  – die Festlegung der Verteilung der Zufallsvariablen  $X$ , d.h. der Wahrscheinlichkeiten

$$\begin{aligned} P(X = x) & \quad \text{für alle } x \in M, \quad \text{bzw. allgemeiner} \\ P(X \in C) & \quad \text{für alle } C \subset M. \end{aligned}$$

Darüber hinaus gehende Konkretisierungen des zu Grunde liegenden W-Raumes  $(\Omega, P)$  und der Abbildung  $X : \Omega \longrightarrow M$  sind nicht notwendig und unterbleiben in der Regel. (Nur vereinzelt und zur Illustration werden wir einen konkreten W-Raum  $(\Omega, P)$  und eine konkrete Abbildung  $X : \Omega \longrightarrow M$  angeben).

**Beispiel 1.2 (Modelle für einige einfache Zufallsexperimente)**

- (a) **Würfel (einmal).**  $X$  beschreibe die gewürfelte Augenzahl, die Ergebnismenge ist  $M = \{1, 2, 3, 4, 5, 6\}$  und es ist  $P(X = x) = 1/6$  für jedes  $x \in M$ . Man nennt  $X$  eine auf der Menge  $\{1, 2, \dots, 6\}$  *gleichverteilte* Zufallsvariable.
- (b) **Würfel (zweimal, unabhängig voneinander).** Ein Würfel wird zweimal geworfen (erster Wurf und zweiter Wurf), oder es werden zwei ‘unterscheidbare’ Würfel (Würfel Nr. 1 und Würfel Nr. 2) gleichzeitig geworfen.

$$M = \{x = (x_1, x_2) : x_1, x_2 \in \{1, 2, 3, 4, 5, 6\}\},$$

$$P(X = x) = 1/36 \quad \text{für jedes } x = (x_1, x_2) \in M.$$

$X$  wird eine auf dieser Menge  $M$  *gleichverteilte* Zufallsvariable genannt. Die Zufallsvariable  $X$  ist hier ‘zwei-dimensional’, da die Werte von  $X$  ja Paare  $x = (x_1, x_2)$  sind. Wir können die zwei-dimensionale Zufallsvariable  $X$  auch als ein Paar ein-dimensionaler Zufallsvariablen auffassen,  $X = (X_1, X_2)$ , d.h.

$$X(\omega) = (X_1(\omega), X_2(\omega)), \quad \text{für } \omega \in \Omega,$$

wobei  $X_1(\omega)$  das Ergebnis des ersten Wurfs und  $X_2(\omega)$  das Ergebnis des zweiten Wurfs angeben; damit schreibt sich die Verteilungsannahme auch als

$$P(X_1 = x_1, X_2 = x_2) = 1/36 \quad \text{für alle } x_1, x_2 \in \{1, 2, 3, 4, 5, 6\},$$

wobei  $P(X_1 = x_1, X_2 = x_2)$  die Wahrscheinlichkeit bezeichnet, dass  $X_1 = x_1$  und  $X_2 = x_2$  eintritt.

- (c) **Lotto-Ziehung 6 aus 49 ohne Zusatzzahl.**

$$M = \{x = (x_1, x_2, x_3, x_4, x_5, x_6) : 1 \leq x_1 < x_2 < x_3 < x_4 < x_5 < x_6 \leq 49, x_i \in \mathbb{N}\},$$

$$P(X = x) = 1 / \binom{49}{6} \quad \text{für jedes } x \in M,$$

wir haben also eine auf dieser Menge  $M$  *gleichverteilte* Zufallsvariable  $X$ . Die Werte von  $X$  sind 6-dimensional, so dass wir auch schreiben können

$$X = (X_1, X_2, X_3, X_4, X_5, X_6),$$

wobei die Zufallsvariablen  $X_1, X_2, \dots, X_6$  die kleinste gezogene Zahl, die zweitkleinste, ..., die größte gezogene Zahl beschreiben.

- (d) **0-1-Experimente (einmalige Durchführung).**

$$M = \{0, 1\}, \quad P(X = 1) = p, \quad P(X = 0) = 1 - p,$$

wobei  $p \in ]0, 1[$  die ‘Erfolgswahrscheinlichkeit’ ist. Interessanter sind *mehrmalige* unabhängige Durchführungen des 0-1-Experiments; das Modell hierfür entwickeln wir später im Zusammenhang mit dem Begriff der *stochastischen Unabhängigkeit* von Zufallsvariablen. In diesem Zusammenhang steht auch das folgende Beispiel einer *binomial-verteilten* zufälligen Anzahl.

- (e) **Eine zufällige Anzahl, binomial-verteilt.** Betrachten wir z.B. das folgende Zufallsexperiment: Ein 0-1-Experiment (mit Erfolgswahrscheinlichkeit  $p \in ]0, 1[$ ) wird 10-mal unabhängig durchgeführt, und als Ergebnis wird die Anzahl der erzielten Erfolge festgehalten. Modell:  $M = \{0, 1, 2, \dots, 10\}$  (klar); gar nicht so klar ist die geeignete W-Verteilung (Begründung später):

$$P(X = x) = \binom{10}{x} p^x (1 - p)^{10-x} \quad \text{für jedes } x = 0, 1, \dots, 10.$$

Man nennt  $X$  eine *binomial-(10, p)-verteilte* Zufallsvariable.

## 1.2 W-Verteilung, W-Raum

### Definition 1.3

Sei  $\Omega$  irgend eine (nicht-leere) Menge. Eine Wahrscheinlichkeitsverteilung auf  $\Omega$ , im Folgenden kurz W-Verteilung auf  $\Omega$  genannt, ist eine Funktion  $P$ , die jeder Teilmenge  $A \subset \Omega$  eine Zahl  $P(A) \in [0, 1]$  als Wahrscheinlichkeit zuordnet und dabei die folgenden beiden Bedingungen erfüllt.

- (i)  $P(\Omega) = 1$ .
- (ii)  $P(A \cup B) = P(A) + P(B)$  für je zwei disjunkte Teilmengen  $A, B \subset \Omega$  (d.h.  $A \cap B = \emptyset$ ).

Wenn  $P$  eine W-Verteilung auf  $\Omega$  ist, dann heißt das Paar  $(\Omega, P)$  ein Wahrscheinlichkeitsraum (kurz: W-Raum).

Im Kontext eines W-Raumes  $(\Omega, P)$  wird jede Teilmenge  $A$  von  $\Omega$  ein Ereignis genannt. Speziell heißt eine ein-elementige Teilmenge  $\{\omega\}$  (für ein einzelnes  $\omega \in \Omega$ ) ein Elementarereignis. (Im Folgenden werden wir das Elementarereignis  $\{\omega\}$  mit dem Element  $\omega$  identifizieren). Ein Ereignis  $A \subset \Omega$ , das aus mehr als nur einem Element besteht, wird auch ein zusammengesetztes Ereignis genannt. Zur Begründung für diese Sprechweisen stellen wir uns vor, dass der Zufall ein Element  $\omega$  aus der Menge  $\Omega$  auswählt. Eine Teilmenge  $A \subset \Omega$  steht dann für das Ereignis, dass dieses zufällig gewählte  $\omega$  zu  $A$  gehört (also  $\omega \in A$  gilt), und  $P(A)$  ist die Wahrscheinlichkeit für dieses Ereignis.

### Beispiel 1.4

Sei  $\Omega$  eine endliche (und nicht-leere) Menge. Wie man leicht sieht, erfüllt dann

$$P(A) = \frac{|A|}{|\Omega|}, \quad A \subset \Omega,$$

die Bedingungen von Definition 1.3 (dabei bezeichnet  $|A|$  die Anzahl der Elemente von  $A$ ). Diese W-Verteilung  $P$  heißt die Gleichverteilung auf  $\Omega$ , weil sie allen Elementarereignissen  $\omega \in \Omega$  die gleiche Wahrscheinlichkeit  $1/|\Omega|$  gibt. ■

Als Folgerungen aus den Bedingungen in Definition 1.3 ergeben sich weitere Rechenregeln für eine W-Verteilung.

### Folgerung 1.5 (Rechenregeln für Wahrscheinlichkeiten)

$$\begin{aligned} P(\bar{A}) &= 1 - P(A), \quad \text{wobei } \bar{A} = \Omega \setminus A, \text{ das Komplement von } A; \\ P(\emptyset) &= 0; \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \quad \text{für beliebige } A, B \subset \Omega; \\ P(A \setminus B) &= P(A) - P(A \cap B) \quad \text{für beliebige } A, B \subset \Omega; \\ \text{aus } A \subset B \subset \Omega &\text{ folgt } P(A) \leq P(B). \end{aligned}$$

Bedingung (ii) der Definition 1.3 nennt man die *Additivität* einer W-Verteilung  $P$ . Es lässt sich folgern, dass dann auch für mehrere paarweise disjunkte Ereignisse eine entsprechende Additivität gilt:



**Folgerung 1.6**

Wenn  $n \in \mathbb{N}$  und  $A_1, A_2, \dots, A_n \subset \Omega$  paarweise disjunkt sind (d.h.  $A_i \cap A_j = \emptyset$  für alle  $i \neq j$ ), dann gilt

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) .$$

Wenn wir *beliebige*, also *nicht* notwendig paarweise disjunkte Ereignisse  $A_1, \dots, A_n \subset \Omega$  hernehmen, dann gilt im Allgemeinen nur

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) .$$

**Folgerung 1.7 (Disjunkte Zerlegung von  $\Omega$ )**

Wenn  $n \in \mathbb{N}$  und  $A_1, \dots, A_n \subset \Omega$  eine *disjunkte Zerlegung von  $\Omega$*  bilden (d.h. die  $A_1, \dots, A_n$  sind paarweise disjunkt mit  $\bigcup_{i=1}^n A_i = \Omega$ ), dann gilt für ein beliebiges Ereignis  $B \subset \Omega$ :

$$B = \bigcup_{i=1}^n (B \cap A_i) .$$

Da dies wiederum eine disjunkte Zerlegung von  $B$  ist, gilt:

$$P(B) = \sum_{i=1}^n P(B \cap A_i) .$$

**Bemerkung 1.8 (Ereignisfeld, Sigma-Additivität)**

Streng mathematisch gesehen ist unsere oben gegebene Definition der Begriffe *W-Verteilung* und *W-Raum* nicht ausreichend. Zum Einen ist es im Fall einer *überabzählbaren* Menge  $\Omega$  (wie z.B.  $\Omega = \mathbb{R}$ ) oft nicht möglich, *alle* Teilmengen von  $\Omega$  als Ereignisse zu betrachten und ihnen in sinnvoller Weise Wahrscheinlichkeiten zuzuordnen. Man muss sich dann auf ein *Teilsystem* des Systems *aller* Teilmengen<sup>2</sup> beschränken, auf ein sog. *Ereignisfeld*. Dieses umfasst aber in der Regel alle praktisch relevanten Teilmengen, so dass die Einschränkung für uns kaum bemerkbar wird. Zum Anderen ist die Eigenschaft der Additivität von  $P$  nicht ausreichend. In der mathematischen Theorie (und an einigen Stellen auch im gegenwärtigen Kontext) wird die etwas weiter gehende Eigenschaft der *Sigma-Additivität* benötigt:

Für jede unendliche Folge paarweise disjunkter Ereignisse  $A_1, A_2, \dots, A_i, \dots \subset \Omega$  (also  $A_i \cap A_j = \emptyset$  für alle  $i, j \in \mathbb{N}$  mit  $i \neq j$ ) gilt

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) .$$

Streng genommen ist diese Eigenschaft anstatt (ii) in Definition 1.3 zu fordern.

## 1.3 Zufallsvariablen

**Definition 1.9**

Seien  $(\Omega, P)$  ein W-Raum und  $M$  eine weitere (nicht-leere) Menge. Dann heißt jede Abbildung  $X : \Omega \longrightarrow M$  eine Zufallsvariable (auf  $\Omega$  mit Werten in  $M$ ).

<sup>2</sup>Das System aller Teilmengen einer Menge  $\Omega$  wird auch *Potenzmenge* von  $\Omega$  genannt.

Der Name *Zufallsvariable* ist so begründet: Wir stellen uns vor, dass die Auswahl eines Elementes  $\omega$  aus der Menge  $\Omega$  *zufällig* geschieht; dann ist auch der Wert  $X(\omega) \in M$ , der unter der Abbildung  $X$  herauskommt, *zufällig*. In diesem Sinne liefert  $X$  zufällige Werte. Die Werte von  $X$  haben Wahrscheinlichkeiten: Für jeden möglichen Wert  $x \in M$  ist die Wahrscheinlichkeit, dass  $X$  gerade diesen Wert  $x$  liefert, gegeben durch  $P(X = x)$ , wobei “ $X = x$ ” für das Ereignis (Teilmenge von  $\Omega$ ) steht,

$$\{X = x\} = \{\omega \in \Omega : X(\omega) = x\} \quad (\subset \Omega),$$

also

$$P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\}).$$

Das erweitern wir auf Ereignisse wie “die Zufallsvariable  $X$  liefert einen Wert, der in  $C$  liegt”, Abk.: “ $X \in C$ ”, für eine Teilmenge  $C \subset M$ ; mathematisch ist dieses Ereignis definiert als das *Urbild* der Menge  $C$  unter der Abbildung  $X$ ,

$$\{X \in C\} = \{\omega \in \Omega : X(\omega) \in C\} \quad (\subset \Omega);$$

die Wahrscheinlichkeit dieses Ereignisses ist

$$P(X \in C) = P(\{\omega \in \Omega : X(\omega) \in C\}).$$

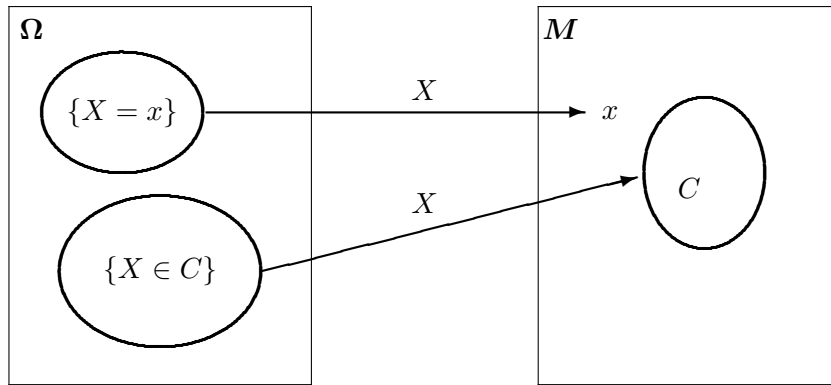


Abbildung 1.1

### Beispiel 1.10

$$\Omega = \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1), (1, 1, 1)\},$$

und die W-Verteilung  $P$  sei die Gleichverteilung auf  $\Omega$  (vergl. Beispiel 1.4). Wir können uns z.B. vorstellen, dass ein Elementarereignis (ein Tripel mit Komponenten 0 oder 1) das Ergebnis dreier unabhängig durchgeführter Münzwürfe darstellt. Interessant sei nun die Anzahl der “1”-en, die ein zufällig ausgewähltes Tripel  $\omega$  aufweist; mathematisch ist damit die folgende Zufallsvariable angesprochen,

$$X : \Omega \longrightarrow M = \{0, 1, 2, 3\}, \quad X(\omega) = \text{Anzahl der “1”en im Tripel } \omega.$$

Ausführlich sieht diese Zufallsvariable so aus:

$\omega$	(0, 0, 0)	(1, 0, 0)	(0, 1, 0)	(0, 0, 1)	(1, 1, 0)	(1, 0, 1)	(0, 1, 1)	(1, 1, 1)
$X(\omega)$	0	1	1	1	2	2	2	3

Betrachten wir z.B. die beiden Ereignisse, dass genau eine “1” im Tripel  $\omega$  auftritt (d.h.  $\{X = 1\}$ ) bzw. dass mindestens zwei “1”en in  $\omega$  auftreten (d.h.  $\{X \in C\}$  mit  $C = \{2, 3\}$ ). Mit Blick auf obige Tabelle ist klar:

$$\begin{aligned}\{X = 1\} &= \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}, \\ \{X \in \{2, 3\}\} &= \{(1, 1, 0), (1, 0, 1), (0, 1, 1), (1, 1, 1)\}.\end{aligned}$$

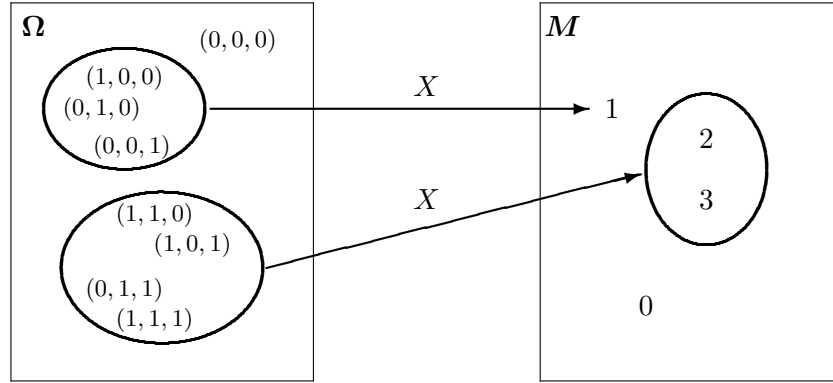


Abbildung 1.2

Die Wahrscheinlichkeiten  $P(X = 1)$  und  $P(X \in \{2, 3\})$  sind – da ja  $P$  die Gleichverteilung auf  $\Omega$  ist – gegeben durch

$$\begin{aligned}P(X = 1) &= \frac{|\{X = 1\}|}{|\Omega|} = \frac{3}{8} = 0.375, \\ P(X \in \{2, 3\}) &= \frac{|\{X \in \{2, 3\}\}|}{|\Omega|} = \frac{4}{8} = 0.5.\end{aligned}$$

■

### Definition 1.11 (Verteilung einer Zufallsvariablen)

Seien  $(\Omega, P)$  ein W-Raum und  $X : \Omega \longrightarrow M$  eine Zufallsvariable (wobei  $M$  irgendeine nicht-leere Menge sei). Die Funktion, die jeder Teilmenge  $C \subset M$  die Wahrscheinlichkeit  $P(X \in C)$  zuordnet, heißt die Verteilung von  $X$  (unter  $P$ ).

Die Verteilung von  $X$  ist eine W-Verteilung auf  $M$ , d.h. die durch

$$Q(C) = P(X \in C) \quad \text{für alle } C \subset M$$

definierte Funktion  $Q$  (welche also jeder Teilmenge  $C \subset M$  die Wahrscheinlichkeit zuordnet, dass die Zufallsvariable  $X$  einen Wert in  $C$  annimmt) hat die in der Definition 1.3 genannten Eigenschaften

- $Q(C) \in [0, 1]$  für alle  $C \subset M$ ;
- (i)  $Q(M) = 1$ ;
- (ii)  $Q$  ist additiv: Wenn  $C, D$  zwei disjunkte Teilmengen von  $M$  sind, dann gilt
$$Q(C \cup D) = Q(C) + Q(D).$$

Das ist nicht schwer zu sehen: Es ist  $P(A) \in [0, 1]$  für jedes  $A \subset \Omega$ , insbesondere also  $P(X \in C) \in [0, 1]$ . Es ist  $Q(M) = P(X \in M) = P(\{\omega \in \Omega : X(\omega) \in M\}) = P(\Omega) = 1$ . Für disjunkte  $C, D \subset \Omega$  sieht man leicht, dass

$$\{X \in C \cup D\} = \{X \in C\} \cup \{X \in D\}$$

und die beiden Mengen  $\{X \in C\}$  und  $\{X \in D\}$  ebenfalls disjunkt sind; daraus ergibt sich mit der Additivität von  $P$ :

$$Q(C \cup D) = P(X \in C \cup D) = P(X \in C) + P(X \in D) = Q(C) + Q(D) .$$

Die Zufallsvariable  $X : \Omega \longrightarrow M$  transformiert sozusagen die W-Verteilung  $P$  (eine W-Verteilung auf  $\Omega$ ) in eine W-Verteilung  $Q$  auf  $M$ , die wir die Verteilung von  $X$  (unter  $P$ ) nennen.

Die Verteilung einer Zufallsvariablen gemäß Definition 1.11 ist im Allgemeinen etwas “unhandlich”, da sie eine Funktion der Variablen  $C \subset M$  ist und als Definitionsbereich also die Menge aller Teilmengen von  $M$  (die sog. *Potenzmenge* von  $M$ ) hat. Im Fall einer *diskreten* Zufallsvariablen  $X$  können wir die Verteilung von  $X$  wesentlich einfacher beschreiben, indem wir uns auf die ein-elementigen Teilmengen  $C = \{x\}$  von  $M$ , d.h. auf die Elemente von  $M$ , zurückziehen und nur die Wahrscheinlichkeiten

$$P(X = x) \quad \text{für alle } x \in M$$

betrachten (das reicht im diskreten Fall aus, s. Theorem 1.13 unten). Wir sprechen von einer *diskreten* Zufallsvariablen, wenn diese nur *abzählbar* viele verschiedene Werte annehmen kann, d.h. wenn ihr Wertevorrat  $M$  als eine *abzählbare* Menge gewählt werden kann. “ $M$  ist *abzählbar*” soll heißen: Entweder ist  $M$  eine *endliche* Menge oder  $M$  ist eine *abzählbar-unendliche* Menge, d.h. die Elemente von  $M$  lassen sich nummerieren,  $x^{(1)}, x^{(2)}, \dots, x^{(i)}, \dots$ , so dass alle Elemente von  $M$  erfasst werden,

$$M = \{x^{(i)} : i \in \mathbb{N}\} .$$

Beispiele für abzählbar-unendliche Mengen sind

$$\begin{aligned} \mathbb{N} &= \{1, 2, 3, \dots\} , \\ \mathbb{N}_0 &= \{0, 1, 2, 3, \dots\} , \\ \mathbb{Z} &= \{0, \pm 1, \pm 2, \pm 3, \dots\} . \end{aligned}$$

Nicht abzählbar (d.h. *überabzählbar*) ist hingegen die Menge  $\mathbb{R}$  aller reellen Zahlen (Zahlengerade) sowie auch jedes nicht-degenerierte Teilintervall von  $\mathbb{R}$ , z.B. abgeschlossene Intervalle  $[a, b]$  mit  $a, b \in \mathbb{R}$ ,  $a < b$ .

### Definition 1.12 (Diskrete Zufallsvariable)

Eine Zufallsvariable  $X : \Omega \longrightarrow M$  mit einer abzählbaren Menge  $M$  heißt eine diskrete Zufallsvariable.

### Theorem 1.13 (Verteilung einer diskreten Zufallsvariablen)

Sei  $X : \Omega \longrightarrow M$  eine diskrete Zufallsvariable. Dann gilt:

$$P(X \in C) = \sum_{x \in C} P(X = x) \quad \text{für alle } C \subset M .$$

Die Verteilung von  $X$  ist daher vollständig beschrieben durch die Wahrscheinlichkeiten

$$P(X = x) \quad \text{für alle } x \in M .$$

Das ist eine Konsequenz aus der Additivitätseigenschaft einer W-Verteilung  $P$  (Bedingung (ii) von Definition 1.3 bzw. Folgerung 1.6, sowie Bemerkung 1.8):

Als Teilmenge der abzählbaren Menge  $M$  ist  $C \subset M$  ebenfalls abzählbar; wenn  $C$  endlich ist und  $x_1, \dots, x_n$  die verschiedenen Elemente von  $C$  bezeichnen, dann ist

$$\{X \in C\} = \bigcup_{i=1}^n \{X = x_i\}, \quad (\text{disjunkte Vereinigung}),$$

folglich

$$P(X \in C) = \sum_{i=1}^n P(X = x_i) = \sum_{x \in C} P(X = x).$$

Wenn  $C$  abzählbar-unendlich ist und  $x_1, \dots, x_i, \dots$  die verschiedenen Elemente von  $C$  in nummerierter Darstellung sind, dann haben wir entsprechend:

$$\{X \in C\} = \bigcup_{i=1}^{\infty} \{X = x_i\}, \quad (\text{disjunkte Vereinigung}),$$

und mit der Sigma-Additivität von  $P$  ist dann

$$P(X \in C) = \sum_{i=1}^{\infty} P(X = x_i) = \sum_{x \in C} P(X = x).$$

### Fortsetzung von Beispiel 1.10

Die Verteilung der Zufallsvariablen  $X$  von Beispiel 1.10 ist beschrieben durch die Wahrscheinlichkeiten

$$P(X = x) = \frac{|\{X = x\}|}{|\Omega|} \quad \text{für } x = 0, 1, 2, 3;$$

aus der Tabelle für  $X$  entnehmen wir:

$$P(X = 0) = \frac{1}{8}, \quad P(X = 1) = \frac{3}{8}, \quad P(X = 2) = \frac{3}{8}, \quad P(X = 3) = \frac{1}{8}.$$

Wir sehen, dass sich diese vier Wahrscheinlichkeiten zu 1 aufsummieren; das ist generell für eine diskrete Zufallsvariable  $X : \Omega \rightarrow M$  der Fall (in unserem Beispiel ist  $M = \{0, 1, 2, 3\}$ ):

$$\sum_{x \in M} P(X = x) = 1.$$

Dies folgt aus Theorem 1.13, wenn wir speziell  $C = M$  wählen:

$$P(X \in M) = \sum_{x \in M} P(X = x),$$

und offensichtlich ist das Ereignis  $\{X \in M\}$  gleich  $\Omega$ , also  $P(X \in M) = P(\Omega) = 1$  (s. Bedingung (i) in Definition 1.3). ■

## 1.4 Wichtige Klassen von diskreten Zufallsvariablen

Mit Hilfe von Theorem 1.13 können wir gewünschte Verteilungen von diskreten Zufallsvariablen definieren: Wir müssen lediglich die Werte für  $P(X = x)$  für alle  $x \in M$  festlegen. Dabei ist nur zu beachten, dass gelten muss:

**Folgerung 1.14**

Die Wahrscheinlichkeiten  $P(X = x)$ ,  $x \in M$ , im Fall einer diskreten Zufallsvariablen  $X : \Omega \longrightarrow M$  erfüllen die Bedingungen:

$$P(X = x) \geq 0 \quad \text{für alle } x \in M, \text{ und}$$

$$\sum_{x \in M} P(X = x) = 1 .$$

Die folgende Definition präsentiert fünf der wichtigsten Verteilungen (genauer: Verteilungsfamilien) für diskrete Zufallsvariablen.

**Definition 1.15**

Seien  $(\Omega, P)$  ein W-Raum,  $M$  eine nicht-leere abzählbare Menge (die sogleich in **(a)** – **(e)** jeweils spezifiziert wird), und sei  $X : \Omega \longrightarrow M$  eine (diskrete) Zufallsvariable.

**(a)** Sei  $M$  endlich. Die Zufallsvariable  $X$  heißt gleichverteilt auf  $M$ , wenn

$$P(X = x) = \frac{1}{|M|} \quad \text{für alle } x \in M.$$

**(b)** Sei  $M = \{0, 1, \dots, n\}$  für ein  $n \in \mathbb{N}$ , und sei  $p \in ]0, 1[$ . Die Zufallsvariable  $X$  heißt binomial- $(n, p)$ -verteilt, wenn

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{für } x = 0, 1, \dots, n .$$

**(c)** Seien  $N \in \mathbb{N}$ ,  $s \in \mathbb{N}_0$  und  $n \in \mathbb{N}$  mit  $s, n \leq N$ , und es sei  $M = \{0, 1, \dots, n\}$ . Die Zufallsvariable  $X$  heißt hypergeometrisch- $(N, s, n)$ -verteilt, wenn

$$P(X = x) = \frac{\binom{s}{x} \binom{N-s}{n-x}}{\binom{N}{n}} \quad \text{für } x = 0, 1, \dots, n .$$

**(d)** Seien  $M = \mathbb{N}_0$  und  $\lambda$  eine positive reelle Zahl. Die Zufallsvariable  $X$  heißt Poisson- $(\lambda)$ -verteilt, wenn

$$P(X = x) = \exp(-\lambda) \cdot \frac{\lambda^x}{x!} \quad \text{für alle } x \in \mathbb{N}_0 .$$

**(e)** Seien  $M = \mathbb{N}$  und  $p \in ]0, 1[$ . Die Zufallsvariable  $X$  heißt geometrisch- $(p)$ -verteilt, wenn

$$P(X = x) = p(1-p)^{x-1} \quad \text{für alle } x \in \mathbb{N} .$$

Wir skizzieren nun typische Zufallsexperimente, deren Modellierung durch Zufallsvariablen gemäß **(a)**–**(e)** von Definition 1.15 erfolgt.

**Typische Zufallsexperimente**

**(a)** Ein Objekt wird “rein zufällig” aus einer endlichen Grundgesamtheit  $M$  gezogen. Die Modellierung erfolgt durch eine Zufallsvariable  $X$  (definiert auf einem W-Raum  $(\Omega, P)$ ) mit Werten in der Menge  $M$ , die auf  $M$  gleichverteilt ist.

- (b)  $n$ -malige unabhängige Durchführung eines 0-1-Experiments mit Wahrscheinlichkeit  $p$  für “1” und  $1 - p$  für “0”, und als Ergebnis des (Gesamt-)Experiments wird die Anzahl der erzielten “1”-en festgehalten. Modellierung dieser zufälligen Anzahl durch eine Zufallsvariable  $X$  mit Werten in  $\{0, 1, \dots, n\}$ , die binomial- $(n, p)$ -verteilt ist.
- (c) Gegeben eine Grundgesamtheit mit  $N$  Objekten; jedes Objekt trägt einen binären (0-1-wertigen) Merkmalswert (z.B. gut/schlecht); insgesamt haben genau  $s$  Objekte den Merkmalswert “1”. Zufallsexperiment: Es wird eine Zufallsstichprobe vom Umfang  $n$  aus der Grundgesamtheit gezogen (Ziehen ohne Zurücklegen), und als Ergebnis wird die Anzahl aller Objekte der Stichprobe mit Merkmalswert “1” festgehalten. Modellierung dieser zufälligen Anzahl durch eine Zufallsvariable  $X$  mit Werten in  $\{0, 1, \dots, n\}$ , die hypergeometrisch- $(N, s, n)$ -verteilt ist.
- (d) Eine zufällige Anzahl, die nicht von vorne herein nach oben beschränkt ist (im Unterschied zu (b) und (c)), kann oft näherungsweise modelliert werden durch eine Zufallsvariable  $X$  mit Werten in  $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ , die Poisson- $(\lambda)$ -verteilt ist (mit einem geeigneten  $\lambda > 0$ ). Konkretere Beispiele etwa: Anzahl von Kunden, die innerhalb eines definierten Zeitraumes eine bestimmte Service-Station besuchen; Anzahl der zerfallenen Teilchen einer radioaktiven Substanz in einem definierten Zeitraum. Mathematische Hinweise auf die Güte der Approximation durch dieses Modell werden wir an späterer Stelle geben (Kapitel 2, Abschnitt 2.3).
- (e) Ein 0-1-Experiment mit Wahrscheinlichkeit  $p$  für “1” und  $1 - p$  für “0” wird so oft unabhängig durchgeführt, bis erstmalig “1” auftritt. Die Anzahl der benötigten Durchführungen ist das Ergebnis des (Gesamt-)Zufallsexperiments. Modellierung dieser zufälligen Anzahl durch eine Zufallsvariable  $X$  mit Werten in  $\mathbb{N} = \{1, 2, \dots\}$ , die geometrisch- $(p)$ -verteilt ist.

**Bemerkung 1.16 (Mathematische Bem.)**

Die in Definition 1.15 (a)–(e) jeweils angesetzten Werte für  $P(X = x)$  (für alle  $x \in M$ ) erfüllen natürlich die notwendigen Bedingungen von Folgerung 1.14 (sonst wäre die Definition ja unsinnig):

Die Nichtnegativität der angegebenen Werte für  $P(X = x)$  ist in allen Fällen (a)–(e) offensichtlich. Die Bedingung, dass die Summe aller Werte  $P(X = x)$  (gebildet über alle  $x \in M$ ) den Wert 1 ergibt, ist in (a) offensichtlich; in (b) folgt sie aus der allgemeinen binomischen Formel

$$(a + b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i} \quad \text{für } a, b \in \mathbb{R}, n \in \mathbb{N}.$$

Wählen wir nämlich speziell  $a = p$  und  $b = 1 - p$ , dann erhalten wir

$$\sum_{i=0}^n \binom{n}{i} p^i (1 - p)^{n-i} = (p + 1 - p)^n = 1.$$

Im Fall (c) erhält man die Summationsbedingung mit Hilfe der kombinatorischen Formel

$$\sum_{i=0}^n \binom{s}{i} \binom{N-s}{n-i} = \binom{N}{n}.$$

In (d) haben wir  $\sum_{x \in \mathbb{N}_0} \exp(-\lambda) \frac{\lambda^x}{x!} = \exp(-\lambda) \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \exp(-\lambda) \exp(\lambda) = 1$ .

In (e) schließlich ist  $\sum_{x \in \mathbb{N}} p(1-p)^{x-1} = p \sum_{i=0}^{\infty} (1-p)^{i-1} = \frac{p}{1 - (1-p)} = 1$ .

## 1.5 Gemeinsame Verteilung mehrerer Zufallsvariablen

Oft besteht ein Zufallsexperiment aus mehreren Einzelexperimenten (z.B. mehrmaliges Würfeln), und dementsprechend ist ein Ergebnis  $x$  des Zufallsexperimentes mehrdimensional,

$$x = (x_1, \dots, x_n),$$

wobei die Komponenten  $x_1, \dots, x_n$  die Ergebnisse der  $n$  Einzelexperimente bezeichnen. Die Modellierung beinhaltet dann entsprechend eine  $n$ -dimensionale Zufallsvariable  $X = (X_1, \dots, X_n)$ . Die Verteilung dieser Zufallsvariablen  $X$  wird auch die *gemeinsame Verteilung der Zufallsvariablen*  $X_1, \dots, X_n$  genannt. Etwas expliziter definieren wir, zunächst für  $n = 2$ :

### Definition 1.17 (Gemeinsame Verteilung zweier ZV'en)

Seien  $(\Omega, P)$  ein W-Raum und  $X_1 : \Omega \rightarrow M_1$  bzw.  $X_2 : \Omega \rightarrow M_2$  zwei Zufallsvariablen auf  $\Omega$  mit Werten in beliebigen Mengen  $M_1$  bzw.  $M_2$ . Dann heißt die Funktion, die jedem Paar  $(C_1, C_2)$  von Teilmengen  $C_1 \subset M_1$  und  $C_2 \subset M_2$  die Wahrscheinlichkeit

$$P(X_1 \in C_1, X_2 \in C_2)$$

zuordnet, die gemeinsame Verteilung von  $X_1$  und  $X_2$ .

Dabei steht wie üblich " $X_1 \in C_1, X_2 \in C_2$ " für das Ereignis

$$\{X_1 \in C_1, X_2 \in C_2\} = \{X_1 \in C_1\} \cap \{X_2 \in C_2\} = \{\omega \in \Omega : X_1(\omega) \in C_1 \text{ und } X_2(\omega) \in C_2\}.$$

Im Fall, dass beide Zufallsvariablen  $X_1$  und  $X_2$  *diskret* sind, d.h. die Mengen  $M_1$  und  $M_2$  abzählbar sind, können wir die gemeinsame Verteilung von  $X_1$  und  $X_2$  einfacher beschreiben, denn dann können wir uns auf die ein-elementigen Teilmengen  $C_1 = \{x_1\}$  und  $C_2 = \{x_2\}$  (mit  $x_1 \in M_1$  und  $x_2 \in M_2$ ) zurückziehen.

### Theorem 1.18 (Gemeinsame Verteilung zweier diskreter ZV'en)

Die gemeinsame Verteilung zweier diskreter Zufallsvariablen  $X_1 : \Omega \rightarrow M_1$  und  $X_2 : \Omega \rightarrow M_2$  ist vollständig beschrieben durch die Wahrscheinlichkeiten

$$P(X_1 = x_1, X_2 = x_2) \quad \text{für alle } x_1 \in M_1 \text{ und } x_2 \in M_2.$$

### Beispiel 1.19 (Zweimaliges unabhängiges Würfeln)

Das Zufallsexperiment des zweimaligen, voneinander unabhängigen Würfeln (s. auch Beispiel 1.2 (b)) wird modelliert durch zwei Zufallsvariablen  $X_1 : \Omega \rightarrow \{1, \dots, 6\}$  und  $X_2 : \Omega \rightarrow \{1, \dots, 6\}$ , deren gemeinsame Verteilung gegeben ist durch

$$P(X_1 = x_1, X_2 = x_2) = \frac{1}{36} \quad \text{für alle } x_1, x_2 \in \{1, \dots, 6\}.$$

Sprechweise:  $X_1$  und  $X_2$  sind *gemeinsam gleichverteilt* auf der Menge der Paare  $\{(x_1, x_2) : x_1, x_2 = 1, \dots, 6\}$ . ■



Die gemeinsame Verteilung zweier Zufallsvariablen  $X_1 : \Omega \longrightarrow M_1$  und  $X_2 : \Omega \longrightarrow M_2$  beinhaltet stets auch die vollständige Information über die Verteilungen der *einzelnen* Zufallsvariablen  $X_1$  und  $X_2$  (die sog. *Randverteilungen* der gemeinsamen Verteilung). Denn wählen wir in Definition 1.17 speziell  $C_2 = M_2$ , dann haben wir (für beliebiges  $C_1 \subset M_1$ )

$$\{X_1 \in C_1, X_2 \in M_2\} = \{X_1 \in C_1\},$$

und die Verteilung von  $X_1$  ist daher zu erhalten gemäß

$$P(X_1 \in C_1) = P(X_1 \in C_1, X_2 \in M_2) \quad \text{für alle } C_1 \subset M_1.$$

Entsprechend erhalten wir die Verteilung von  $X_2$  durch

$$P(X_2 \in C_2) = P(X_1 \in M_1, X_2 \in C_2) \quad \text{für alle } C_2 \subset M_2.$$

Im Fall zweier *diskreter* Zufallsvariablen  $X_1$  und  $X_2$  haben wir expliziter:

**Theorem 1.20 (Randverteilungen im Fall zweier diskreter ZV'en)**

Für zwei diskrete Zufallsvariablen  $X_1 : \Omega \longrightarrow M_1$  und  $X_2 : \Omega \longrightarrow M_2$  gilt;

$$\begin{aligned} P(X_1 = x_1) &= \sum_{x_2 \in M_2} P(X_1 = x_1, X_2 = x_2) \quad \text{für alle } x_1 \in M_1, \text{ und} \\ P(X_2 = x_2) &= \sum_{x_1 \in M_1} P(X_1 = x_1, X_2 = x_2) \quad \text{für alle } x_2 \in M_2. \end{aligned}$$

Dies ist eine Konsequenz aus Folgerung 1.7:

Sei etwa  $M_2$  endlich,  $M_2 = \{x_2^{(1)}, \dots, x_2^{(n)}\}$ . Dann ist durch  $A_i = \{X_2 = x_2^{(i)}\}$  ( $i = 1, \dots, n$ ) eine disjunkte Zerlegung von  $\Omega$  gegeben, und daher – mit  $B = \{X_1 = x_1\}$  für ein  $x_1 \in M_1$  – nach Folgerung 1.7:

$$P(X_1 = x_1) = P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(X_1 = x_1, X_2 = x_2^{(i)}) = \sum_{x_2 \in M_2} P(X_1 = x_1, X_2 = x_2).$$

Entsprechend behandelt man den Fall, dass die Menge  $M_2$  *abzählbar-unendlich* ist,  $M_2 = \{x_2^{(i)} : i \in \mathbb{N}\}$ ; dann treten *unendliche* Summen auf, aber die Argumentation bleibt dieselbe (wobei hier die *Sigma*-Additivität von  $P$  zum Tragen kommt).

Die *gemeinsame* Verteilung zweier Zufallsvariablen beinhaltet also insbesondere die *einzelnen* Verteilungen der Zufallsvariablen (*Randverteilungen*). Aber: Die gemeinsame Verteilung geht darüber hinaus und beinhaltet auch eventuell vorhandene Abhängigkeiten zwischen den beiden Zufallsvariablen. Im Allgemeinen lässt sich die gemeinsame Verteilung nicht allein aus den einzelnen Verteilungen der beiden Zufallsvariablen bestimmen (die wesentliche Ausnahme ist der Fall zweier *stochastisch unabhängiger* Zufallsvariablen, s. Kapitel 2, Abschnitt 2.2).

**Beispiel 1.21 (Zweimal Ziehen mit/ohne Zurücklegen)**

Aus einer Grundgesamtheit von  $N$  Objekten ( $N \in \mathbb{N}$ ,  $N \geq 2$ ) werden zufällig nacheinander zwei Objekte herausgezogen. Der Einfachheit halber stellen wir uns vor, die Objekte seien durchgehend nummeriert mit  $1, \dots, N$ . Wir unterscheiden zwei Varianten dieses Zufallsexperimentes:

- (a) Es wird *mit Zurücklegen* gezogen, d.h. das zuerst gezogene Objekt wird vor dem zweiten Zug der Grundgesamtheit wieder zugeführt.
- (b) Es wird *ohne Zurücklegen* gezogen.

Modellierung für Variante (a):

Sie erfolgt durch die beiden Zufallsvariablen  $X_1$  (Nummer des zuerst gezogenen Objektes) und  $X_2$  (Nummer des als zweites gezogenen Objektes). Die Ergebnismenge von  $X_1$  und  $X_2$  ist jeweils  $\{1, \dots, N\}$ . Für die gemeinsame Verteilung von  $X_1$  und  $X_2$  haben wir anzusetzen (wobei wir hier schreiben:  $i$  statt  $x_1$  und  $j$  statt  $x_2$ ),

$$P(X_1 = i, X_2 = j) = \frac{1}{N^2} \quad \text{für alle } i, j \in \{1, \dots, N\}.$$

Wir berechnen nun die Verteilungen der *einzelnen* Zufallsvariablen mit Hilfe von Theorem 1.18:

$$P(X_1 = i) = \sum_{j=1}^N \underbrace{P(X_1 = i, X_2 = j)}_{=1/N^2} = N \frac{1}{N^2} = \frac{1}{N}$$

für jedes  $i \in \{1, \dots, N\}$ . Also ist  $X_1$  eine auf der Menge  $\{1, \dots, N\}$  *gleichverteilte* Zufallsvariable. Analoge Rechnung ergibt dies auch für  $X_2$ .

Modellierung für Variante (b):

Sie erfolgt durch die beiden Zufallsvariablen  $Y_1$  (Nummer des zuerst gezogenen Objektes) und  $Y_2$  (Nummer des als zweites gezogenen Objektes). Die Ergebnismenge von  $Y_1$  und  $Y_2$  ist jeweils  $\{1, \dots, N\}$ . Für die gemeinsame Verteilung von  $Y_1$  und  $Y_2$  haben wir anzusetzen:

$$P(Y_1 = i, Y_2 = j) = \begin{cases} \frac{1}{N(N-1)} & , \text{ falls } i \neq j \\ 0 & , \text{ falls } i = j \end{cases}, \quad \text{für alle } i, j \in \{1, \dots, N\}.$$

Für die Randverteilungen ergibt sich mit Theorem 1.18:

$$P(Y_1 = i) = \sum_{j=1}^N P(Y_1 = i, Y_2 = j) = \sum_{j: j \neq i} \underbrace{P(Y_1 = i, Y_2 = j)}_{=1/(N(N-1))} = (N-1) \frac{1}{N(N-1)} = \frac{1}{N}$$

für jedes  $i \in \{1, \dots, N\}$ . Also ist auch  $Y_1$  eine auf  $\{1, \dots, N\}$  *gleichverteilte* Zufallsvariable, und analog erhalten wir dies auch für  $Y_2$ .

Wir sehen:

Die *einzelnen* Verteilungen der Zufallsvariablen sind in beiden Fällen (a) und (b) dieselben, aber die *gemeinsamen* Verteilungen in (a) und (b) sind verschieden. ■

Wir dehnen noch die Betrachtungen auf den Fall einer beliebigen (endlichen) Anzahl  $n \geq 2$  von Zufallsvariablen aus:

Seien  $X_i : \Omega \longrightarrow M_i$  ( $i = 1, \dots, n$ ) Zufallsvariablen, wobei  $(\Omega, P)$  ein W-Raum und  $M_1, \dots, M_n$  beliebige (nicht-leere) Mengen sind.

**Definition 1.22 (Gemeinsame Verteilung von  $n$  ZV'en)**

Unter der gemeinsamen Verteilung der Zufallsvariablen  $X_1, \dots, X_n$  versteht man die Funktion, die jedem  $n$ -Tupel  $(C_1, \dots, C_n)$  von Teilmengen  $C_i \subset M_i$ , ( $i = 1, \dots, n$ ), die Wahrscheinlichkeit

$$P(X_1 \in C_1, X_2 \in C_2, \dots, X_n \in C_n)$$

zuordnet.

Dabei steht " $X_1 \in C_1, X_2 \in C_2, \dots, X_n \in C_n$ " für das Ereignis

$$\begin{aligned} \{X_1 \in C_1, X_2 \in C_2, \dots, X_n \in C_n\} &= \{X_1 \in C_1\} \cap \{X_2 \in C_2\} \cap \dots \cap \{X_n \in C_n\} \\ &= \{\omega \in \Omega : X_i(\omega) \in C_i \text{ für alle } i = 1, \dots, n\}. \end{aligned}$$

Im Fall, dass die Zufallsvariablen  $X_1, \dots, X_n$  *diskret* sind, d.h. die Mengen  $M_1, \dots, M_n$  abzählbar sind, können wir die gemeinsame Verteilung von  $X_1, \dots, X_n$  einfacher beschreiben:

**Theorem 1.23 (Gemeinsame Verteilung von  $n$  diskreten ZV'en)**

Die gemeinsame Verteilung von  $n$  diskreten Zufallsvariablen  $X_i : \Omega \longrightarrow M_i$ ,  $i = 1, \dots, n$ , ist vollständig beschrieben durch die Wahrscheinlichkeiten

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \quad \text{für alle } x_i \in M_i, i = 1, \dots, n.$$

Die gemeinsame Verteilung von  $n$  Zufallsvariablen  $X_i : \Omega \longrightarrow M_i$ ,  $i = 1, \dots, n$ , beinhaltet stets auch die Verteilungen der *einzelnen* Zufallsvariablen  $X_1, \dots, X_n$  (die sog. *Randverteilungen* der gemeinsamen Verteilung), denn betrachten wir etwa die Zufallsvariable  $X_1$ :

Wählen wir in Definition 1.22 speziell  $C_j = M_j$  für alle  $j = 2, \dots, n$ , dann haben wir (für beliebiges  $C_1 \subset M_1$ )

$$\{X_1 \in C_1, X_2 \in M_2, \dots, X_n \in M_n\} = \{X_1 \in C_1\},$$

und die Verteilung von  $X_1$  ist daher zu erhalten gemäß

$$P(X_1 \in C_1) = P(X_1 \in C_1, X_2 \in M_2, \dots, X_n \in M_n) \quad \text{für alle } C_1 \subset M_1.$$

Entsprechend erhalten wir auch die Verteilung von  $X_i$  für jedes  $i = 2, \dots, n$ .

Für den Fall *diskreter* Zufallsvariablen  $X_1, \dots, X_n$  lässt sich die Formel von Theorem 1.20 ebenfalls erweitern, worauf wir aber verzichten wollen; beispielhaft formulieren wir nur eine Formel im Fall dreier diskreter Zufallsvariablen  $X_1, X_2, X_3$ :

$$P(X_1 = x_1) = \sum_{x_2 \in M_2} \sum_{x_3 \in M_3} P(X_1 = x_1, X_2 = x_2, X_3 = x_3) \quad \text{für alle } x_1 \in M_1.$$

## Kapitel 2

# Stochastische Unabhängigkeit, bedingte Wahrscheinlichkeiten

### 2.1 Stochastische Unabhängigkeit von Ereignissen

Im Folgenden gehen wir von einem W-Raum  $(\Omega, P)$  aus. Der Begriff der stochastischen Unabhängigkeit von Ereignissen in  $\Omega$  bezieht sich auf die W-Verteilung  $P$ . Betrachten wir diesen Begriff zunächst für zwei Ereignisse.

**Definition 2.1 (Stoch. Unabhängigkeit zweier Ereignisse)**

Zwei Ereignisse  $A, B \subset \Omega$  heißen stochastisch unabhängig, wenn

$$P(A \cap B) = P(A) \cdot P(B) .$$

**Beispiel 2.2**

Sei  $\Omega$  die Menge aller Paare  $\omega = (\omega_1, \omega_2)$  mit  $\omega_1, \omega_2 \in \{1, 2, 3, 4, 5, 6\}$ , und sei  $P$  die Gleichverteilung auf  $\Omega$ . Die beiden Ereignisse

$$\begin{aligned} A &= \{ \omega \in \Omega : \omega_1 = 6 \} = \{ (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6) \} , \\ B &= \{ \omega \in \Omega : \omega_2 \in \{5, 6\} \} = \{ (1, 5), (2, 5), \dots, (6, 5), (1, 6), (2, 6), \dots, (6, 6) \} \end{aligned}$$

sind stochastisch unabhängig, denn:

$$A \cap B = \{ \omega \in \Omega : \omega_1 = 6, \omega_2 \in \{5, 6\} \} = \{ (6, 5), (6, 6) \} ,$$

also  $P(A \cap B) = \frac{2}{36} = \frac{1}{18}$ ; andererseits  $P(A) \cdot P(B) = \frac{6}{36} \cdot \frac{12}{36} = \frac{1}{6} \cdot \frac{1}{3} = \frac{1}{18}$ .

Hingegen sind die beiden folgenden Ereignisse nicht stochastisch unabhängig, d.h. sie sind stochastisch abhängig.

$$\begin{aligned} A &= \{ \omega \in \Omega : \omega_1 = 6 \text{ oder } \omega_2 = 6 \} = \{ (6, 1), (6, 2), \dots, (6, 6), (1, 6), (2, 6), \dots, (5, 6) \} , \\ B &= \{ \omega \in \Omega : \omega_2 \in \{2, 3\} \} = \{ (1, 2), (2, 2), \dots, (6, 2), (1, 3), (2, 3), \dots, (6, 3) \} . \end{aligned}$$

Dann ist nämlich

$$A \cap B = \{ (6, 2), (6, 3) \} ,$$

also  $P(A \cap B) = \frac{2}{36} = \frac{1}{18}$ ; andererseits:  $P(A) \cdot P(B) = \frac{11}{36} \cdot \frac{12}{36} = \frac{11}{108} < \frac{1}{18}$ . ■

Definition 2.1 wird nun erweitert auf mehrere ( $n \geq 2$ ) Ereignisse in  $\Omega$ :

**Definition 2.3 (Stoch. Unabhängigkeit von  $n$  Ereignissen)**

Seien  $n \in \mathbb{N}$ ,  $n \geq 2$ , und  $A_1, \dots, A_n \subset \Omega$ . Die Ereignisse  $A_1, \dots, A_n$  heißen stochastisch unabhängig, wenn gilt:

Für jedes  $k = 2, \dots, n$  und für jede Auswahl von  $k$  Indizes  $i_1, i_2, \dots, i_k$  aus  $1, \dots, n$  mit  $i_1 < \dots < i_k$  ist

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_k}) .$$

Im Fall  $n = 2$  ist dies offensichtlich dasselbe wie Definition 2.1 (nur mit anderer Notation:  $A_1, A_2$  statt  $A, B$ ). Im Fall  $n = 3$  sagt Definition 2.3: Die Ereignisse  $A_1, A_2, A_3$  sind genau dann stochastisch unabhängig, wenn die folgenden vier Gleichungen erfüllt sind.

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1) P(A_2) , \\ P(A_1 \cap A_3) &= P(A_1) P(A_3) , \\ P(A_2 \cap A_3) &= P(A_2) P(A_3) , \\ P(A_1 \cap A_2 \cap A_3) &= P(A_1) P(A_2) P(A_3) . \end{aligned}$$

Stochastische Unabhängigkeit bedeutet also mehr als nur paarweise stochastische Unabhängigkeit.

## 2.2 Stochastische Unabhängigkeit von Zufallsvariablen

Wir gehen wieder von einem W-Raum  $(\Omega, P)$  aus. Der Begriff der *stochastischen Unabhängigkeit* von Zufallsvariablen auf  $\Omega$  bezieht sich wiederum auf die W-Verteilung  $P$ . Beginnen wir mit dem Fall zweier Zufallsvariablen auf  $\Omega$ .

**Definition 2.4 (Stoch. Unabhängigkeit zweier ZV'en)**

Seien  $X_1 : \Omega \rightarrow M_1$  und  $X_2 : \Omega \rightarrow M_2$  zwei Zufallsvariablen auf  $\Omega$ , wobei  $M_1$  und  $M_2$  zwei (nicht-leere) Mengen seien. Die Zufallsvariablen  $X_1$  und  $X_2$  heißen stochastisch unabhängig, wenn für jede Auswahl von Teilmengen  $C_1 \subset M_1$  und  $C_2 \subset M_2$  gilt:

Die beiden Ereignisse  $\{X_1 \in C_1\}$  und  $\{X_2 \in C_2\}$  sind stochastisch unabhängig, d.h.

$$P(X_1 \in C_1, X_2 \in C_2) = P(X_1 \in C_1) \cdot P(X_2 \in C_2) .$$

Mit Blick auf Definition 1.17, lässt sich das auch so formulieren:

*Die beiden Zufallsvariablen  $X_1$  und  $X_2$  sind genau dann stochastisch unabhängig, wenn ihre gemeinsame Verteilung gleich dem Produkt ihrer einzelnen Verteilungen ist.*

Die Erweiterung von Definition 2.4 auf mehrere Zufallsvariablen geschieht in nahe liegender Weise durch Rückführung auf Definition 2.3 (stochastische Unabhängigkeit mehrerer Ereignisse).

**Definition 2.5 (Stoch. Unabhängigkeit von  $n$  ZV'en)**

Seien  $X_1, X_2, \dots, X_n$  Zufallsvariablen auf  $\Omega$ , also  $X_i : \Omega \rightarrow M_i$ , für  $i = 1, 2, \dots, n$  (wobei  $M_1, M_2, \dots, M_n$  beliebige nicht-leere Mengen seien). Die Zufallsvariablen  $X_1, X_2, \dots, X_n$  heißen stochastisch unabhängig, wenn für jede Auswahl von Teilmengen  $C_1 \subset M_1, C_2 \subset M_2, \dots, C_n \subset M_n$  gilt:

Die Ereignisse  $\{X_1 \in C_1\}, \{X_2 \in C_2\}, \dots, \{X_n \in C_n\}$  sind stochastisch unabhängig.

Etwas “formelhafter” ausgedrückt:

Die Zufallsvariablen  $X_1, X_2, \dots, X_n$  sind genau dann stochastisch unabhängig, wenn für jede Auswahl von Teilmengen  $C_1 \subset M_1, C_2 \subset M_2, \dots, C_n \subset M_n$  gilt

$$P(X_1 \in C_1, X_2 \in C_2, \dots, X_n \in C_n) = P(X_1 \in C_1) \cdot P(X_2 \in C_2) \cdot \dots \cdot P(X_n \in C_n). \quad (2.1)$$

Mit Blick auf Definition 1.22, können wir das auch so ausdrücken:

Die Zufallsvariablen  $X_1, \dots, X_n$  sind genau dann stochastisch unabhängig, wenn ihre gemeinsame Verteilung gleich dem Produkt ihrer einzelnen Verteilungen ist.

### Bemerkung 2.6

Die Gültigkeit von Gleichung (2.1) für sich genommen, d.h. für feste  $C_1, C_2, \dots, C_n$  gelesen, ist im Allgemeinen nicht dasselbe wie die stochastische Unabhängigkeit der Ereignisse  $\{X_1 \in C_1\}, \{X_2 \in C_2\}, \dots, \{X_n \in C_n\}$ . Denn Letztere besagt ja nach Definition 2.3, dass

$$P(X_{i_1} \in C_{i_1}, \dots, X_{i_k} \in C_{i_k}) = P(X_{i_1} \in C_{i_1}) \cdot \dots \cdot P(X_{i_k} \in C_{i_k}) \quad (2.2)$$

für jedes  $k = 2, \dots, n$  und jede Auswahl  $i_1 < \dots < i_k$  von  $k$  Indizes aus  $1, 2, \dots, n$ . Hingegen steht in (2.1) nur der Fall  $k = n$ . Aber: Mit dem Zusatz, dass die Gültigkeit von (2.1) verlangt wird für jede Auswahl der  $C_i \subset M_i, i = 1, \dots, n$ , ist diese Bedingung in der Tat äquivalent mit der stochastischen Unabhängigkeit der Zufallsvariablen  $X_1, X_2, \dots, X_n$  gemäß Definition 2.5; (um (2.2) aus (2.1) zu erhalten, wähle  $C_i = M_i$  für alle Indizes  $i$ , die nicht unter den  $i_1 < \dots < i_k$  vorkommen). ■

Im Fall diskreter Zufallsvariablen läßt sich die Beschreibung der stochastischen Unabhängigkeit wesentlich vereinfachen:

### Theorem 2.7 (Stoch. Unabhängigkeit für diskrete ZV'en)

Seien  $X_1, X_2, \dots, X_n$  diskrete Zufallsvariablen auf  $\Omega$ , d.h.  $X_i : \Omega \longrightarrow M_i$  mit abzählbaren (nicht-leeren) Mengen  $M_i, (i = 1, \dots, n)$ . Die Zufallsvariablen  $X_1, X_2, \dots, X_n$  sind genau dann stochastisch unabhängig, wenn für jede Auswahl von Elementen  $x_1 \in M_1, x_2 \in M_2, \dots, x_n \in M_n$  gilt

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \dots \cdot P(X_n = x_n).$$

Spezieller, für den Fall zweier ( $n = 2$ ) diskreter Zufallsvariablen  $X_1$  und  $X_2$ , lautet die Bedingung für die stochastische Unabhängigkeit von  $X_1$  und  $X_2$  also:

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2) \quad \text{für alle } x_1 \in M_1 \text{ und } x_2 \in M_2.$$

### Beispiel 2.8 (vergl. Beispiel 1.3 (b))

Seien  $X_1$  und  $X_2$  zwei Zufallsvariablen auf  $\Omega$  jeweils mit Werten in der Menge  $\{1, \dots, 6\}$ , und  $X_1, X_2$  seien gemeinsam gleichverteilt auf der Menge aller Paare der Zahlen  $1, \dots, 6$ , d.h.

$$P(X_1 = i, X_2 = j) = 1/36 \quad \text{für alle } i, j \in \{1, 2, 3, 4, 5, 6\}.$$

Wir prüfen, ob  $X_1$  und  $X_2$  stochastisch unabhängig sind (mit Theorem 2.7). Dazu müssen wir  $P(X_1 = i)$  für jedes  $i = 1, 2, \dots, 6$  berechnen, und auch  $P(X_2 = j)$  für jedes  $j = 1, 2, \dots, 6$  (also die *Randverteilungen*). Das können wir leicht mit Hilfe von Theorem 1.20:

$$P(X_1 = i) = \sum_{j=1}^6 \underbrace{P(X_1 = i, X_2 = j)}_{=1/36} = \frac{6}{36} = \frac{1}{6}.$$

Analog erhalten wir für jedes  $j = 1, 2, \dots, 6$ :

$$P(X_2 = j) = \frac{1}{6}.$$

Damit ist

$$P(X_1 = i) \cdot P(X_2 = j) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = P(X_1 = i, X_2 = j);$$

die beiden Zufallsvariablen  $X_1, X_2$  sind also stochastisch unabhängig. Insgesamt hat sich herausgestellt: Die *gemeinsame* Gleichverteilung der Zufallsvariablen  $X_1, X_2$  (auf der Menge der Paare der Zahlen  $1, \dots, 6$ ) impliziert die Gleichverteilung sowohl von  $X_1$  als auch von  $X_2$  auf der Menge  $\{1, 2, \dots, 6\}$  sowie die stochastische Unabhängigkeit von  $X_1$  und  $X_2$ .

**Beispiel 2.9 (2-mal Ziehen mit/ohne Zurücklegen, vergl. Beispiel 1.21)**

Aus einer Grundgesamtheit von  $N$  Objekten ( $N \in \mathbb{N}$ ,  $N \geq 2$ ) werden zufällig, nacheinander zwei Objekte herausgezogen. Wir unterscheiden die Varianten:

- (a) Ziehen mit Zurücklegen;
- (b) Ziehen ohne Zurücklegen.

Wie in Beispiel 1.21 dargelegt, erfolgt die Modellierung so:

Im Fall (a): Durch zwei Zufallsvariablen  $X_1, X_2$  auf  $\Omega$  mit Werten jeweils in  $\{1, \dots, N\}$ , die *gemeinsam* gleichverteilt sind (auf der Menge der Paare der Zahlen  $1, \dots, N$ ), d.h.

$$P(X_1 = i, X_2 = j) = \frac{1}{N^2} \quad \text{für alle } i, j = 1, \dots, N.$$

Im Fall (b): Durch zwei Zufallsvariablen  $Y_1, Y_2$  auf  $\Omega$  mit Werten jeweils in  $\{1, \dots, N\}$ , deren gemeinsame Verteilung gegeben ist durch

$$P(Y_1 = i, Y_2 = j) = \begin{cases} \frac{1}{N(N-1)} & , \text{ falls } i \neq j \\ 0 & , \text{ falls } i = j \end{cases} \quad , \quad \text{für alle } i, j = 1, \dots, N.$$

In Beispiel 1.21 haben wir gesehen, dass jede der Zufallsvariablen  $X_1, X_2, Y_1, Y_2$  einzeln gleichverteilt auf der Menge  $\{1, \dots, N\}$  ist. Daraus folgt:

Im Fall (a): Die beiden Zufallsvariablen  $X_1$  und  $X_2$  sind stochastisch unabhängig, da

$$\underbrace{P(X_1 = i, X_2 = j)}_{=1/N^2} = \underbrace{P(X_1 = i)}_{=1/N} \cdot \underbrace{P(X_2 = i)}_{=1/N}.$$

Im Fall (b): Die beiden Zufallsvariablen  $Y_1$  und  $Y_2$  sind nicht stochastisch unabhängig (also stochastisch abhängig), da

$$P(Y_1 = i) \cdot P(Y_2 = j) = \frac{1}{N} \cdot \frac{1}{N} = \frac{1}{N^2},$$

was offensichtlich verschieden von  $P(Y_1 = i, Y_2 = j)$  ist (letzteres ist ja gleich  $\frac{1}{N(N-1)}$ , falls  $i \neq j$ , und gleich 0, falls  $i = j$ ).

## 2.3 Modellierung unabhängiger Zufallsexperimente

Oft haben wir es mit Zufallsexperimenten zu tun, die aus mehreren “unabhängigen” Einzelexperimenten bestehen (z.B. die mehrmalige unabhängige Durchführung eines 0-1-Experiments). Mit Hilfe des Begriffes der stochastischen Unabhängigkeit lässt sich eine geeignete Modellierung dann allgemein wie folgt vornehmen.

Nehmen wir an, wir haben ein Zufallsexperiment, das aus  $n$  unabhängigen Einzelexperimenten besteht; ein Ergebnis des (gesamten) Zufallsexperiments besteht in der Auflistung aller Einzelergebnisse, ist also mathematisch ausgedrückt ein  $n$ -Tupel

$$x = (x_1, x_2, \dots, x_n),$$

wobei  $x_i$  das Ergebnis des  $i$ -ten Einzelexperimentes bezeichnet,  $i = 1, 2, \dots, n$ . Zur Modellierung gehen wir von einem W-Raum  $(\Omega, P)$  und Zufallsvariablen  $X_1, \dots, X_n$  auf  $\Omega$  aus, mit der Interpretation, dass ein  $\omega \in \Omega$  zufällig (entsprechend  $P$ ) realisiert wird und die Werte  $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$  der Zufallsvariablen die Ergebnisse der Einzelexperimente liefern. Für die einzelnen Verteilungen der Zufallsvariablen  $X_i$  sind dabei natürlich geeignete Annahmen zu treffen ( $X_i$  soll ja das  $i$ -te Einzelexperiment adäquat beschreiben). Die Unabhängigkeit der Einzelexperimente wird als stochastische Unabhängigkeit der Zufallsvariablen  $X_1, X_2, \dots, X_n$  in die Modellierung eingebracht.

Das Modell für das Gesamtexperiment sieht daher folgendermaßen aus:

Ein Ergebnis  $x = (x_1, x_2, \dots, x_n)$  des (Gesamt-)Zufallsexperiments besteht aus Werten der stochastisch unabhängigen Zufallsvariablen  $X_1, X_2, \dots, X_n$ , wobei  $X_i$  das  $i$ -te Einzelexperiment beschreibt (für  $i = 1, 2, \dots, n$ ). Die Verteilung der möglichen Ergebnisse  $x$  ist daher durch die *gemeinsame* Verteilung von  $X_1, \dots, X_n$  gegeben, die wegen der stochastischen Unabhängigkeit gleich dem *Produkt der einzelnen* Verteilungen ist:

$$P(X_1 \in C_1, X_2 \in C_2, \dots, X_n \in C_n) = P(X_1 \in C_1) \cdot P(X_2 \in C_2) \cdot \dots \cdot P(X_n \in C_n)$$

für jede Wahl von Teilmengen  $C_1 \subset M_1, C_2 \subset M_2, \dots, C_n \subset M_n$  (wobei  $M_1, \dots, M_n$  die Mengen der möglichen Ergebnisse der  $n$  Einzelexperimente bezeichnen).

Im Fall diskreter Zufallsvariablen  $X_1, \dots, X_n$  (d.h. abzählbarer Mengen  $M_1, \dots, M_n$ ) vereinfacht sich dieses Formelsystem zu:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \dots \cdot P(X_n = x_n)$$

für jede Auswahl von Elementen  $x_1 \in M_1, x_2 \in M_2, \dots, x_n \in M_n$ .

### Beispiel 2.10 ( $n$ -malige unabhängige Durchführung eines 0-1-Experiments)

Die Wahrscheinlichkeit für “1” im Einzelexperiment sei mit  $p \in ]0, 1[$  bezeichnet. Das Modell für das Gesamtexperiment ( $n$ -malige unabhängige Durchführung des 0-1-Experiments) ist gegeben durch stochastisch unabhängige 0-1-wertige Zufallsvariablen  $X_1, X_2, \dots, X_n$  mit

$$P(X_i = 1) = p \quad \text{und} \quad P(X_i = 0) = 1 - p, \quad i = 1, 2, \dots, n.$$

Die Wahrscheinlichkeiten im Gesamtexperiment für die einzelnen  $n$ -Tupel  $x = (x_1, x_2, \dots, x_n)$  (wobei die  $x_i \in \{0, 1\}$ ) sind daher gegeben durch

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) &= P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \dots \cdot P(X_n = x_n) \\ &= p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}. \end{aligned}$$

Hier ist  $\sum_{i=1}^n x_i$  gleich der Anzahl der erzielten “1”-en und  $n - \sum_{i=1}^n x_i$  die Anzahl der erzielten “0”-en.



**Beispiel 2.11 (Binomial-verteilte Anzahl)**

In der Praxis wird man bei einem Zufallsexperiment wie in Beispiel 2.10 nicht die ausführliche Sequenz ( $n$ -Tupel)  $x = (x_1, x_2, \dots, x_n)$  als Ergebnis dokumentieren, sondern die Reduktion auf die Anzahl der “1”-en vornehmen. Das Ergebnis des Zufallsexperiments ist dann ein Wert einer Zufallsvariablen  $S$  (von *Summe*, sagen wir), die ihre möglichen Werte in der Menge  $\{0, 1, \dots, n\}$  hat. Die adäquate Verteilungsannahme ist dann:

*$S$  ist eine binomial- $(n, p)$ -verteilte Zufallsvariable.*

(Vergl. Definition 1.15 (b) sowie “Typische Zufallsexperimente” Punkt (b); vergl. auch Beispiel 1.3 (e).)

Dies lässt sich so einsehen:

Mit den 0-1-wertigen Zufallsvariablen  $X_1, \dots, X_n$  haben wir

$$S = \sum_{i=1}^n X_i, \quad \text{d.h. } S(\omega) = \sum_{i=1}^n X_i(\omega) \quad \text{für alle } \omega \in \Omega.$$

Damit können wir die Verteilung von  $S$ , d.h. die Wahrscheinlichkeiten  $P(S = k)$  für alle  $k = 0, 1, \dots, n$ , berechnen (unter Verwendung des Ergebnisses von Beispiel 2.10): Das Ereignis  $\{S = k\}$  ist gleich der Vereinigung der paarweise disjunkten Ereignisse

$$\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\},$$

genommen über alle  $n$ -Tupel  $x = (x_1, x_2, \dots, x_n)$  mit  $\sum_{i=1}^n x_i = k$ . Folglich ist  $P(S = k)$  die Summe aller Wahrscheinlichkeiten dieser Ereignisse. Für jedes  $x = (x_1, x_2, \dots, x_n)$  mit  $\sum_{i=1}^n x_i = k$  erhalten wir aber denselben Wert (s. Beispiel 2.10),

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = p^k (1-p)^{n-k},$$

so dass  $P(S = k)$  gleich diesem Wert mal der Anzahl der  $n$ -Tupel  $x$  mit  $\sum_{i=1}^n x_i = k$  ist. Die Anzahl dieser  $x$  ist gleich  $\binom{n}{k}$ . Wir haben also:

$$P(S = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{für } k = 0, 1, 2, \dots, n.$$

Die Zufallsvariable  $S$  ist daher binomial- $(n, p)$ -verteilt. ■

Schwieriger ist die Modellierung einer zufälligen Anzahl (von “1”-en, etwa “Erfolgen”) in  $n$  unabhängigen 0-1-Experimenten, wenn die einzelnen “Erfolgswahrscheinlichkeiten”  $p_1, \dots, p_n$  nicht notwendig alle identisch sind. Mathematisch ist damit die Verteilung einer Summenvariablen  $S = \sum_{i=1}^n X_i$  angesprochen, wobei  $X_1, \dots, X_n$  stochastisch unabhängige 0-1-Variablen sind, mit

$$P(X_i = 1) = p_i, \quad P(X_i = 0) = 1 - p_i, \quad i = 1, \dots, n,$$

und  $p_i \in ]0, 1[$  für  $i = 1, \dots, n$ . Hier liefert eine Poisson-Verteilung (Definition 1.15 (d)), eine brauchbare Approximation, sofern die Wahrscheinlichkeiten  $p_i$  klein sind (eine Poisson-Verteilung wird daher auch eine *Verteilung für seltene Ereignisse* genannt). Daher ist oft plausibel, dass eine zufällige Anzahl, z.B. die Anzahl der Besucher einer Service-Station an einem Tag, oder z. B. die Anzahl zerfallener Teilchen einer radioaktiven Substanz in einer Sekunde, in guter Näherung durch eine Poisson-verteilte Zufallsvariable beschreibbar ist. Man kann sich nämlich vorstellen, dass die zufällige Anzahl durch eine Reihe von unabhängigen 0-1-Experimenten mit jeweils kleinen Wahrscheinlichkeiten für “1” zu Stande kommt; so wird im Beispiel der Service-Station ein größerer Personenkreis von potentiellen Kunden existieren, wobei jede Person mit einer gewissen kleinen Wahrscheinlichkeit die Service-Station an dem betrachteten Tag tatsächlich aufsucht. (Vergl. auch Abschnitt 1.4, “Typische Zufallsexperimente”, Punkt (d)).

**Bemerkung 2.12 (Poisson-Verteilung)**

Für eine zufällige Anzahl  $S$ , die eine Summe von  $n$  stochastisch unabhängigen 0-1-Variablen  $X_1, X_2, \dots, X_n$  ist, lässt sich ihre Verteilung approximativ durch eine Poisson- $(\lambda)$ -Verteilung mit  $\lambda = \sum_{i=1}^n p_i$  beschreiben, sofern die Wahrscheinlichkeiten

$$p_i = P(X_i = 1) \quad \text{für } i = 1, 2, \dots, n$$

klein (aber positiv) sind.

Für die Abweichung zwischen der exakten Verteilung von  $S$  und der Verteilung einer Poisson- $(\lambda)$ -verteilten Zufallsvariablen  $X$  (mit obigem  $\lambda$ ) lässt sich die folgende Abschätzung herleiten:

$$\left| P(S \in C) - P(X \in C) \right| \leq \sum_{i=1}^n p_i^2 \quad \text{für alle } C \subset \mathbb{N}_0.$$

Auch im Fall *identischer* Wahrscheinlichkeiten  $p_1 = \dots = p_n = p$ , wie in Beispiel 2.11, ist die approximative Beschreibung einer binomial- $(n, p)$ -verteilten Zufallsvariablen durch eine Poisson- $(\lambda = np)$ -verteilte Zufallsvariable bisweilen nützlich, nämlich dann, wenn  $n$  groß und  $p$  klein sind. Die obere Schranke für die Abweichung zwischen den beiden Verteilungen (binomial- $(n, p)$  und Poisson- $(\lambda = np)$ ) ist dann  $np^2$ . Z.B. für  $n = 1000$  und  $p = 0.005$  ergibt sich  $np^2 = 0.025$ , d.h. der Verteilungsunterschied zwischen einer binomial- $(n = 1000, p = 0.005)$ -verteilten Zufallsvariablen  $S$  und einer Poisson- $(\lambda = 5)$ -verteilten Zufallsvariablen  $X$  ist höchstens 0.025.

**Beispiel 2.13 (Warten auf den ersten Erfolg: Geometrische Verteilung)**

Ein 0-1-Experiment werde so oft unabhängig durchgeführt, bis erstmalig "1" auftritt. Die benötigte Anzahl  $n$  der Durchführungen ist hier das Ergebnis des (Gesamt-)Zufallsexperiments. Zur Modellierung nehmen wir daher als Ergebnismenge

$$M = \mathbb{N} \cup \{\infty\},$$

wobei wir zunächst auch  $\infty$  als ein mögliches Ergebnis zulassen, da es denkbar ist, dass die Folge der 0-1-Experimente niemals eine "1" liefert (also jede Durchführung "0" als Ergebnis bringt). Zur weiteren Modellierung gehen wir wie üblich von einem W-Raum  $(\Omega, P)$  und einer Zufallsvariablen  $W : \Omega \rightarrow M$  aus, die das zufällige Ergebnis liefert (die Bezeichnung  $W$  hier von *Wartezeit*). Desweiteren sei  $X_1, X_2, \dots, X_i, \dots$  eine unendliche Folge 0-1-wertiger Zufallsvariablen auf  $\Omega$ , welche die Ergebnisse der einzelnen 0-1-Experimente beschreiben. Die Wahrscheinlichkeit für "1" im 0-1-Experiment sei mit  $p \in ]0, 1[$  bezeichnet. Also haben wir

$$P(X_i = 1) = p, \quad P(X_i = 0) = 1 - p \quad \text{für jedes } i = 1, 2, \dots$$

Die einzelnen Durchführungen des 0-1-Experiments sollen voneinander unabhängig erfolgen, so dass wir die Zufallsvariablen  $X_i$  ( $i = 1, 2, \dots$ ) als stochastisch unabhängig voraussetzen. Da es sich hier um unendlich viele Zufallsvariablen handelt, bedarf dies einer Erläuterung: Die stochastische Unabhängigkeit der (unendlichen) Folge von Zufallsvariablen  $X_i$ , ( $i = 1, 2, \dots$ ), bedeutet, dass für jedes  $n \in \mathbb{N}$ ,  $n \geq 2$ , die Zufallsvariablen  $X_1, X_2, \dots, X_n$  stochastisch unabhängig sind.

Die Zufallsvariable  $W$  hängt mit den 0-1-Variablen so zusammen:

$$W(\omega) = \min\{i \in \mathbb{N} : X_i(\omega) = 1\} \quad \text{für alle } \omega \in \Omega,$$

wobei wir unter dem Minimum der leeren Menge den Wert  $\infty$  verstehen wollen. Jetzt können wir die Verteilung der Zufallsvariablen  $W$  bestimmen. Für ein  $n \in \mathbb{N}$  erhalten wir:

Im Fall  $n = 1$ :

$$P(W = 1) = P(X_1 = 1) = p;$$

im Fall  $n \geq 2$ :

$$\begin{aligned} P(W = n) &= P(X_1 = 0, \dots, X_{n-1} = 0, X_n = 1) \\ &= P(X_1 = 0) \cdot \dots \cdot P(X_{n-1} = 0) \cdot P(X_n = 1) = (1-p)^{n-1} p. \end{aligned}$$

Beide Fälle für  $n$  lassen sich zusammenfassen zu

$$P(W = n) = p(1-p)^{n-1} \quad \text{für alle } n \in \mathbb{N}.$$

Es verbleibt noch die Berechnung von  $P(W = \infty)$ . Das tun wir mit Hilfe des komplementären Ereignisses  $\{W \in \mathbb{N}\}$ :

$$\begin{aligned} P(W = \infty) &= 1 - P(W \in \mathbb{N}) = 1 - \sum_{n=1}^{\infty} P(W = n) \\ &= 1 - \sum_{n=1}^{\infty} p(1-p)^{n-1} = 1 - p \underbrace{\sum_{n=1}^{\infty} (1-p)^{n-1}}_{=1/(1-(1-p))} = 1 - \frac{p}{p} = 0. \end{aligned}$$

Die Wahrscheinlichkeit für das Ereignis  $\{W = \infty\}$  ist also gleich Null, so dass wir den Wert  $\infty$  aus der Ergebnismenge  $M$  ruhig entfernen können (was auch der Erfahrung entspricht, dass in der Folge der unabhängig durchgeführten 0-1-Experimente irgendwann eine "1" auftreten wird). Mit Blick auf Definition 1.15 (e) stellen wir fest:

*Die Zufallsvariable  $W$  (Wartezeit auf die erste "1") ist geometrisch-( $p$ )-verteilt.*

(Vergl. auch "Typische Zufallsexperimente", Punkt (e).)

## 2.4 Bedingte Wahrscheinlichkeiten

### Definition 2.14

Seien  $(\Omega, P)$  ein W-Raum und  $A, B \subset \Omega$  mit  $P(B) > 0$ . Dann heißt die Zahl

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

die bedingte Wahrscheinlichkeit von  $A$  unter  $B$ .

Es gilt  $0 \leq P(A|B) \leq 1$ , denn:  $P(A|B) \geq 0$  ist offensichtlich;  $P(A|B) \leq 1$  folgt aus  $A \cap B \subset B$ , also  $P(A \cap B) \leq P(B)$  (s. Abschnitt 1.2).

Die Interpretation des Wertes  $P(A|B)$  ist folgende: Im Unterschied zur "unbedingten" Wahrscheinlichkeit  $P(A)$  des Ereignisses  $A$  ist  $P(A|B)$  die aktualisierte Wahrscheinlichkeit von  $A$  unter der Zusatzinformation, dass das Ereignis  $B$  eintritt (oder eingetreten ist).

Wann bleibt die Wahrscheinlichkeit für  $A$  dieselbe, d.h. wann gilt  $P(A|B) = P(A)$ ? Das ist offensichtlich genau dann der Fall, wenn  $P(A \cap B) = P(A) \cdot P(B)$  ist, d.h. wenn die Ereignisse  $A$  und  $B$  stochastisch unabhängig sind.

### Beispiel 2.15

Sei  $\Omega = \{\omega = (\omega_1, \omega_2) : \omega_1, \omega_2 \in \{1, 2, 3, 4, 5, 6\}\}$  und  $P$  die Gleichverteilung auf  $\Omega$ . Betrachte das Ereignis

$$A = \{\omega = (\omega_1, \omega_2) \in \Omega : \omega_1 = 6 \text{ oder } \omega_2 = 6\}.$$

Auflistung aller Elemente von  $A$  ergibt  $|A| = 11$ , folglich  $P(A) = 11/36 \approx 0.306$ . Jetzt sei die Zusatzinformation gegeben, dass das Ereignis  $\{\omega_1 + \omega_2 \geq 8\}$  eingetreten ist, d.h. das Ereignis

$$B = \{ \omega = (\omega_1, \omega_2) \in \Omega : \omega_1 + \omega_2 \geq 8 \} .$$

Die Auflistung der Elemente von  $B$  und von  $A \cap B$  ergibt  $|B| = 15$ , also  $P(B) = 15/36$ , und  $|A \cap B| = 9$ , also  $P(A \cap B) = 9/36$ . Wir erhalten

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{9/36}{15/36} = \frac{9}{15} = 0.6 .$$

Durch die Zusatzinformation, dass das Ereignis  $B$  eingetreten ist, hat sich also die Wahrscheinlichkeit für das Ereignis  $A$  von  $P(A) \approx 0.306$  auf  $P(A|B) = 0.6$  erhöht.

### Bemerkung 2.16

Wenn wir die bedingte Wahrscheinlichkeit  $P(A|B)$  für variierendes  $A \subset \Omega$  betrachten, aber mit einem festen  $B \subset \Omega$  (mit  $P(B) > 0$ ), so sehen wir unschwer die folgenden Eigenschaften:

$$\begin{aligned} P(\Omega|B) &= 1 \quad \text{und auch} \quad P(B|B) = 1; \\ P(A_1 \cup A_2|B) &= P(A_1|B) + P(A_2|B) \quad \text{für je zwei disjunkte } A_1, A_2 \subset \Omega. \end{aligned}$$

Gemäß Definition 1.4 ist daher  $P(A|B)$  als Funktion von  $A \subset \Omega$  (bei festem  $B$ ) eine W-Verteilung auf  $\Omega$ ; diese ist sozusagen die Aktualisierung der W-Verteilung  $P$  unter der Zusatzinformation, dass das Ereignis  $B$  eintritt (oder eingetreten ist). Die Eigenschaft  $P(B|B) = 1$  ist daher ebenfalls einleuchtend. ■

Direkt aus Definition 2.14 erhalten wir die Formel (für Ereignisse  $A, B \subset \Omega$  mit  $P(B) > 0$ ),

$$P(A \cap B) = P(A|B) \cdot P(B) . \quad (2.3)$$

Diese Formel können wir in eine weitergehende Formel für eine beliebige Anzahl  $n \geq 2$  von Ereignissen einordnen. Dazu schreiben wir (2.3) in etwas anderer Notation:

$$P(A_1 \cap A_2) = P(A_1) P(A_2|A_1) , \quad \text{für } A_1, A_2 \subset \Omega \text{ mit } P(A_1) > 0. \quad (2.4)$$

Für drei Ereignisse  $A_1, A_2, A_3 \subset \Omega$  ist

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \frac{P(A_1 \cap A_2)}{P(A_1)} \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} = P(A_1) P(A_2|A_1) P(A_3|A_1 \cap A_2) ,$$

wobei allerdings vorauszusetzen ist, dass die jeweiligen bedingenden Ereignisse positive Wahrscheinlichkeiten haben (da sonst die bedingte Wahrscheinlichkeit nicht definiert wäre). Also ist dabei vorauszusetzen, dass  $P(A_1) > 0$  und  $P(A_1 \cap A_2) > 0$ , was sich wegen der allgemeinen Gültigkeit von  $P(A_1 \cap A_2) \leq P(A_1)$  zur Voraussetzung  $P(A_1 \cap A_2) > 0$  reduzieren lässt. Damit haben wir für drei Ereignisse  $A_1, A_2, A_3 \subset \Omega$  mit  $P(A_1 \cap A_2) > 0$  die Formel:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) P(A_2|A_1) P(A_3|A_1 \cap A_2) \quad (2.5)$$

Allgemeiner gilt für eine beliebige Anzahl  $n \geq 2$  von Ereignissen:

### Theorem 2.17 (Multiplikationssatz für bedingte Wahrscheinlichkeiten)

Wenn  $n \in \mathbb{N}$ ,  $n \geq 2$ , und  $A_1, A_2, \dots, A_n$  Ereignisse in  $\Omega$  sind mit  $P\left(\bigcap_{j=1}^{n-1} A_j\right) > 0$ , dann gilt:

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1) \prod_{i=2}^n P\left(A_i \left| \bigcap_{j=1}^{i-1} A_j\right.\right) .$$

**Beispiel 2.18 ( $n$ -maliges Ziehen ohne Zurücklegen: "gut/schlecht"-Prüfung)**

Eine Grundgesamtheit bestehe aus  $N$  Objekten, die jeweils ein 0-1-wertiges Merkmal tragen (z.B. gut/schlecht); insgesamt seien  $s$  Objekte mit dem Merkmalswert 1 vorhanden. Es werden  $n$  Objekte zufällig, nacheinander und ohne Zurücklegen gezogen und ihre Merkmalswerte festgestellt. Die Ergebnisse der einzelnen Ziehungen sind (Werte von) Zufallsvariablen  $X_1, X_2, \dots, X_n$ , wobei  $X_i : \Omega \rightarrow \{0, 1\}$  für  $i = 1, 2, \dots, n$ , mit einem W-Raum  $(\Omega, P)$ . Zur Beschreibung des *gesamten* Zufallsexperiments bestimmen wir die Wahrscheinlichkeiten

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \quad \text{für alle } x_1, x_2, \dots, x_n \in \{0, 1\}.$$

Diese können wir berechnen unter den plausiblen Annahmen, dass gilt:

$$P(X_1 = 1) = \frac{s}{N}, \quad P(X_1 = 0) = \frac{N-s}{N}, \quad (2.6)$$

$$P(X_i = 1 \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}) = \frac{s - \sum_{j=1}^{i-1} x_j}{N - (i-1)} \quad \text{für } i = 2, \dots, n, \quad (2.7)$$

$$P(X_i = 0 \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}) = \frac{N - s - \sum_{j=1}^{i-1} (1 - x_j)}{N - (i-1)} \quad \text{für } i = 2, \dots, n.$$

Anwendung von Theorem 2.17 auf die Ereignisse  $A_i = \{X_i = x_i\}$  (für  $i = 1, 2, \dots, n$ ) liefert

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_1) \prod_{i=2}^n P(X_i = x_i \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \quad (2.8)$$

Betrachten wir z.B. den Fall  $n = 5$ ,  $x_1 = 0$ ,  $x_2 = 1$ ,  $x_3 = 1$ ,  $x_4 = 0$ ,  $x_5 = 0$ . Wir erhalten mit (2.6), (2.7) und (2.8):

$$\begin{aligned} P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 0) &= \frac{N-s}{N} \frac{s}{N-1} \frac{s-1}{N-2} \frac{N-s-1}{N-3} \frac{N-s-2}{N-4} \\ &= \frac{s(s-1)(N-s)(N-s-1)(N-s-2)}{N(N-1)(N-2)(N-3)(N-4)}. \end{aligned}$$

Für allgemeine  $n$  und  $x = (x_1, \dots, x_n)$  erhalten wir die folgende Formel, wobei wir zur Abkürzung setzen  $k(x) = \sum_{i=1}^n x_i$  und  $\bar{k}(x) = n - k(x)$ .

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) &= \frac{s(s-1) \dots (s - [k(x) - 1]) (N-s)(N-s-1) \dots (N-s - [\bar{k}(x) - 1])}{N(N-1) \dots (N - (n-1))} \\ &= \frac{\binom{s}{k(x)} \binom{N-s}{n-k(x)}}{\binom{N}{n} \binom{n}{k(x)}}, \quad \text{wobei } k(x) = \sum_{i=1}^n x_i \text{ und } \bar{k}(x) = n - k(x). \end{aligned} \quad (2.9)$$

**Beispiel 2.19 (Hypergeometrisch-verteilte Anzahl)**

In der Praxis wird man bei einem Zufallsexperiment wie in Beispiel 2.18 nicht die ausführliche Sequenz ( $n$ -Tupel)  $x = (x_1, x_2, \dots, x_n)$  als Ergebnis dokumentieren, sondern die Reduktion auf die Anzahl der "1"-en vornehmen. Das Ergebnis des Zufallsexperiments ist dann ein Wert einer Zufallsvariablen  $T$ , die ihre möglichen Werte in der Menge  $\{0, 1, \dots, n\}$  hat. Die adäquate Verteilungsannahme ist dann:

*$T$  ist eine hypergeometrisch- $(N, s, n)$ -verteilte Zufallsvariable.*

(Vergl. Definition 1.15 (c) sowie "Typische Zufallsexperimente" Punkt (c)).

Dies lässt sich so einsehen:

Mit den 0-1-wertigen Zufallsvariablen  $X_1, \dots, X_n$  haben wir

$$T = \sum_{i=1}^n X_i, \quad \text{d.h. } T(\omega) = \sum_{i=1}^n X_i(\omega) \quad \text{für alle } \omega \in \Omega.$$

Damit können wir die Verteilung von  $T$ , d.h. die Wahrscheinlichkeiten  $P(T = k)$  für alle  $k = 0, 1, \dots, n$  berechnen (unter Verwendung des Ergebnisses von Beispiel 2.18): Das Ereignis  $\{T = k\}$  ist gleich der Vereinigung der paarweise disjunkten Ereignisse

$$\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\},$$

genommen über alle  $n$ -Tupel  $x = (x_1, x_2, \dots, x_n)$  mit  $k(x) = k$ . Folglich ist  $P(T = k)$  die Summe der Wahrscheinlichkeiten dieser Ereignisse. Für jedes  $x = (x_1, x_2, \dots, x_n)$  mit  $k(x) = k$  erhalten wir aber denselben Wert (s. obige Formel (2.9)),

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{\binom{s}{k} \binom{N-s}{n-k}}{\binom{N}{n} \binom{n}{k}},$$

so dass  $P(T = k)$  gleich diesem Wert mal der Anzahl der  $n$ -Tupel  $x$  mit  $k(x) = k$  ist. Die Anzahl dieser  $x$  ist gleich  $\binom{n}{k}$ . Wir haben also:

$$P(T = k) = \binom{n}{k} \frac{\binom{s}{k} \binom{N-s}{n-k}}{\binom{N}{n} \binom{n}{k}} = \frac{\binom{s}{k} \binom{N-s}{n-k}}{\binom{N}{n}} \quad \text{für } k = 0, 1, 2, \dots, n.$$

Die Zufallsvariable  $T$  ist daher hypergeometrisch- $(N, s, n)$ -verteilt. ■

Zwei weitere Formeln im Zusammenhang mit bedingten Wahrscheinlichkeiten präsentiert das folgende Theorem.

### Theorem 2.20 (Bayes'sche Formeln)

Seien  $n \in \mathbb{N}$  und  $A_1, A_2, \dots, A_n$  eine disjunkte Zerlegung von  $\Omega$  mit  $P(A_i) > 0$  für alle  $i = 1, 2, \dots, n$ . Dann gilt für jedes Ereignis  $B \subset \Omega$ :

$$P(B) = \sum_{i=1}^n P(B|A_i) P(A_i). \quad (\text{"Formel von der totalen Wahrscheinlichkeit"})$$

Wenn  $P(B) > 0$  ist, dann gilt noch:

$$P(A_j|B) = \frac{P(B|A_j) P(A_j)}{\sum_{i=1}^n P(B|A_i) P(A_i)} \quad \text{für jedes } j = 1, 2, \dots, n.$$

Diese Formeln sind leicht einzusehen:

Da  $A_1, A_2, \dots, A_n$  die Menge  $\Omega$  disjunkt zerlegen, ist nach Folgerung 1.7

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B|A_i) P(A_i).$$

Das ist die erste Formel. In der zweiten Formel ist der Nenner auf der rechten Seite daher gleich  $P(B)$ , und die Formel ist gleichbedeutend mit

$$P(A_j|B) = \frac{P(B|A_j) P(A_j)}{P(B)},$$

und diese Gleichung ist richtig, was man anhand Definition 2.14 der bedingten Wahrscheinlichkeiten sofort verifiziert.

**Beispiel 2.21**

Ein Hersteller von Computern bezieht ein bestimmtes Bauteil von drei Zulieferern. Die folgenden Anteilswerte an der Gesamtlieferung sowie die jeweiligen Ausschussanteile innerhalb der drei Lieferungen sind aufgrund längerer Erfahrung bekannt.

	Zulieferer		
	1	2	3
Anteil an Ges.	60 %	25 %	15 %
Ausschussanteil in Lief.	8 %	12 %	4 %

Wir fassen die relativen Häufigkeiten als Wahrscheinlichkeiten auf:

Betrachte die Menge  $\Omega$  aller gelieferten Bauteile, sei  $P$  die Gleichverteilung auf  $\Omega$ .

Dann haben wir für die Teilmengen von  $\Omega$ ,

$A_i$  – die von Zulieferer  $i$  stammenden Bauteile,  $i = 1, 2, 3$ ;

$B$  – die Ausschuss-Bauteile,

die folgenden Wahrscheinlichkeiten und bedingten Wahrscheinlichkeiten:

$$P(A_1) = 0.6, \quad P(A_2) = 0.25, \quad P(A_3) = 0.15, \\ P(B|A_1) = 0.08, \quad P(B|A_2) = 0.12, \quad P(B|A_3) = 0.04.$$

- (a) Wie groß ist der Ausschussanteil in der Gesamtlieferung?

Mit der ersten Bayes'schen Formel:

$$P(B) = \sum_{i=1}^3 P(B|A_i) P(A_i) = 0.08 \cdot 0.6 + 0.12 \cdot 0.25 + 0.04 \cdot 0.15 = 0.084.$$

Der Ausschussanteil an der Gesamtlieferung beträgt also 8.4 %.

- (b) Welche Anteile am Gesamtausschuss haben die einzelnen Zulieferer?

Die zweite Bayes'sche Formel ergibt:

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{\sum_{i=1}^3 P(B|A_i)P(A_i)} = \frac{P(B|A_1)P(A_1)}{P(B)} = \frac{0.08 \cdot 0.6}{0.084} \approx 0.57;$$

ein Ausschussteil stammt also mit 57%-iger Wahrscheinlichkeit von Zulieferer 1.

Analog erhalten wir für Zulieferer 2 und Zulieferer 3:

$$P(A_2|B) = \frac{0.12 \cdot 0.25}{0.084} \approx 0.36, \\ P(A_3|B) = \frac{0.04 \cdot 0.15}{0.084} \approx 0.07.$$

Die Anteile der Zulieferer 2 und 3 am Gesamtausschuss betragen also 36 % bzw. 7 %.

# Kapitel 3

## Verteilungen reeller Zufallsvariablen

### 3.1 Verteilungsfunktionen

Sei  $(\Omega, P)$  ein W-Raum. Eine Zufallsvariable auf  $\Omega$  mit Werten in der Menge  $\mathbb{R}$  der reellen Zahlen, also

$$X : \Omega \longrightarrow \mathbb{R} ,$$

heißt eine reelle Zufallsvariable. Die Verteilung der reellen Zufallsvariablen  $X$  ist gemäß der allgemeinen Definition 1.11 gegeben durch die Wahrscheinlichkeiten

$$P(X \in C) \quad \text{für alle } C \subset \mathbb{R} . \quad (3.1)$$

Das ist allerdings eine sehr unhandliche Beschreibung, da das System aller Teilmengen  $C$  von  $\mathbb{R}$  (die sog. Potenzmenge von  $\mathbb{R}$ ) sehr groß und unübersichtlich ist. Andererseits ist es im Allgemeinen nicht ausreichend, in (3.1) nur die ein-elementigen Teilmengen  $C = \{x\}$  von  $\mathbb{R}$  und ihre Wahrscheinlichkeiten,

$$P(X = x) \quad \text{für alle } x \in \mathbb{R} ,$$

zu betrachten. Dies ist ein wesentlicher Unterschied zum Fall einer diskreten Zufallsvariablen (vergl. Kap. 1, Abschnitt 1.5; siehe auch Abschnitt 3.2 unten). Für eine nicht-diskrete reelle Zufallsvariable  $X$  haben wir typischerweise

$$P(X = x) = 0 \quad \text{für alle } x \in \mathbb{R} ,$$

(vergl. Bemerkung 3.3 unten), so dass diese Wahrscheinlichkeiten wenig über die Verteilung der Zufallsvariablen aussagen können.

Glücklicherweise müssen wir dennoch nicht alle möglichen Teilmengen  $C \subset \mathbb{R}$  in (3.1) zur Beschreibung der Verteilung einer reellen Zufallsvariablen  $X$  betrachten. Auch interessieren in der Regel “exotische” Teilmengen  $C \subset \mathbb{R}$  nicht, sondern in erster Linie Intervalle  $C = [a, b]$  (für  $a, b \in \mathbb{R}$  mit  $a < b$ ), oder auch  $C = ]a, b[$ ,  $C = ]a, b]$  oder unbeschränkte Intervalle  $C = ]-\infty, b]$ ,  $C = [a, \infty[$  usw. Für die entsprechenden Wahrscheinlichkeiten verwendet man die folgenden Schreibweisen:

$$\begin{aligned} P(X \in [a, b]) &= P(a \leq X \leq b) , \\ P(X \in ]a, b[) &= P(a < X < b) , \\ P(X \in ]a, b]) &= P(a < X \leq b) , \\ P(X \in ]-\infty, b]) &= P(X \leq b) , \\ P(X \in [a, \infty[) &= P(X \geq a) . \end{aligned}$$



Ein zentrales mathematisches Resultat in diesem Zusammenhang besagt, dass die Verteilung einer reellen Zufallsvariablen  $X$  bereits vollständig beschrieben ist durch die Wahrscheinlichkeiten für alle Intervalle der Form  $C = ] - \infty, b]$ , wobei  $b \in \mathbb{R}$ , d.h. durch die Wahrscheinlichkeiten

$$P(X \leq b) \quad \text{für alle } b \in \mathbb{R}.$$

Diese Wahrscheinlichkeiten, als eine Funktion von  $b$  betrachtet, bilden die Verteilungsfunktion der Zufallsvariablen  $X$ . Den üblichen Notationen folgend ersetzen wir noch das Symbol  $b$  durch das Symbol  $x$  und formulieren die folgende Definition.

**Definition 3.1 (Verteilungsfunktion einer reellen ZV'en)**

Seien  $(\Omega, P)$  ein W-Raum und  $X : \Omega \rightarrow \mathbb{R}$  eine reelle Zufallsvariable auf  $\Omega$ . Dann heißt die Funktion

$$F(x) = P(X \leq x), \quad \text{für alle } x \in \mathbb{R},$$

die Verteilungsfunktion der Zufallsvariablen  $X$ .

Offensichtlich gilt  $0 \leq F(x) \leq 1$  für alle  $x \in \mathbb{R}$ . Weitergehende Eigenschaften der Verteilungsfunktion einer reellen Zufallsvariablen  $X$  sind im folgenden Theorem festgehalten.

**Theorem 3.2 (Eigenschaften der Verteilungsfunktion)**

Für die Verteilungsfunktion  $F$  einer reellen Zufallsvariablen  $X$  gilt:

- (i)  $F$  ist monoton wachsend, d.h. für alle  $x_1, x_2 \in \mathbb{R}$  mit  $x_1 \leq x_2$  ist  $F(x_1) \leq F(x_2)$ .
- (ii)  $F$  ist rechtsseitig stetig, d.h. für jedes  $x \in \mathbb{R}$  ist

$$\lim_{z \rightarrow x, z > x} F(z) = F(x).$$

- (iii)  $\lim_{x \rightarrow \infty} F(x) = 1$  und  $\lim_{x \rightarrow -\infty} F(x) = 0$ .

- (iv) Für die linksseitigen Grenzwerte von  $F$ , d.h.  $F_\ell(x) := \lim_{z \rightarrow x, z < x} F(z)$ , ist

$$F(x) - F_\ell(x) = P(X = x) \quad \text{für jedes } x \in \mathbb{R}.$$

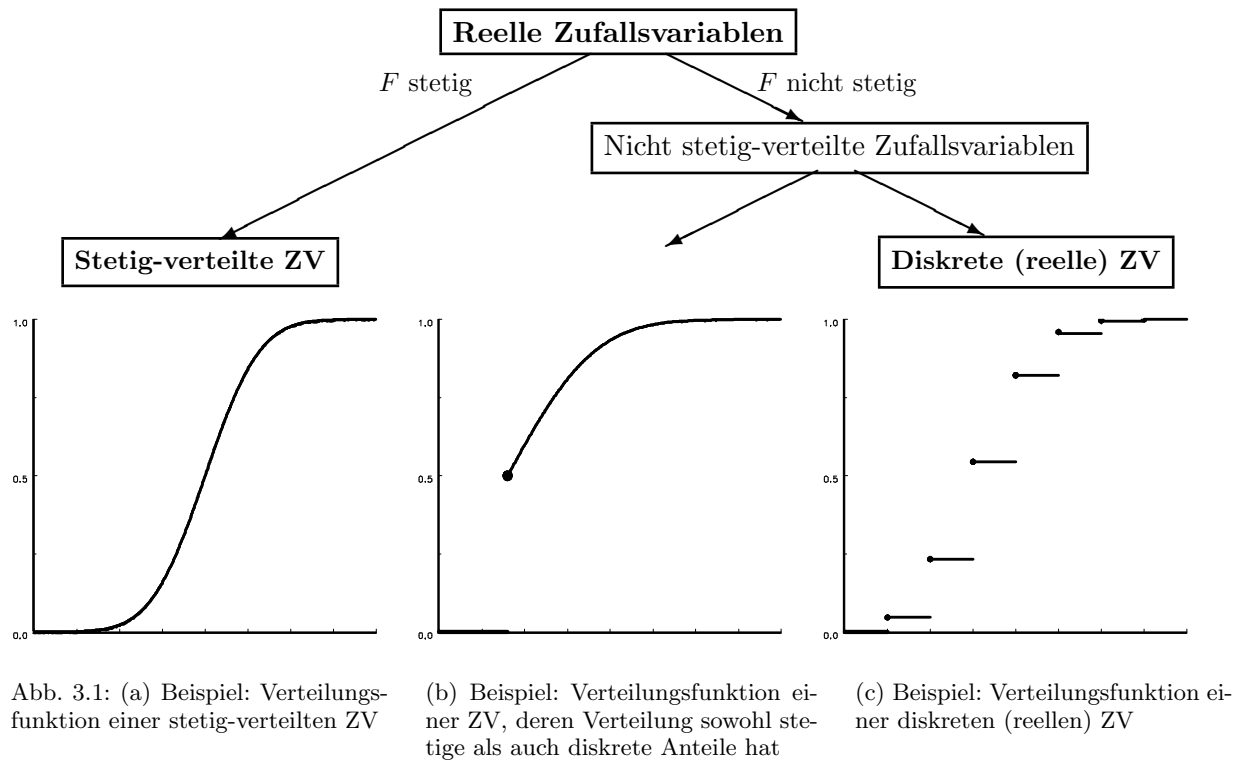
- (v) 
$$\begin{aligned} P(a < X \leq b) &= F(b) - F(a) \quad \text{für alle } a, b \in \mathbb{R} \text{ mit } a < b, \\ P(X > a) &= 1 - F(a) \quad \text{für alle } a \in \mathbb{R}. \end{aligned}$$

**Bemerkung 3.3 (Stetig-verteilte ZV'e)**

Wenn die Verteilungsfunktion der Zufallsvariablen  $X$  stetig ist, wie z.B. in Abbildung 3.1 (a), so sprechen wir von einer stetig-verteilten Zufallsvariablen. Nach Theorem 3.2, Punkte (ii) und (iv), ist die Verteilungsfunktion von  $X$  genau dann stetig, wenn gilt

$$P(X = x) = 0 \quad \text{für alle } x \in \mathbb{R}.$$

Stetig-verteilte Zufallsvariablen betrachten wir in Abschnitt 3.3. Aber die Klasse der reellen Zufallsvariablen schließt als Spezialfall auch diskrete (reelle) Zufallsvariablen mit ein, wie der folgende Abschnitt zeigt.



### 3.2 Diskrete Zufallsvariablen

Wir haben schon in Kap. 1 spezielle reelle Zufallsvariablen betrachtet (z.B. in Definition 1.15 (b)-(e)): Dies waren *diskrete* Zufallsvariablen, da ihre möglichen Werte tatsächlich in einer abzählbaren Teilmenge  $M \subset \mathbb{R}$  lagen, (z.B.  $M = \{0, 1, \dots, n\}$ ,  $M = \mathbb{N}_0$  oder  $M = \mathbb{N}$ ).

Eine Zufallsvariable

$$X : \Omega \longrightarrow M ,$$

mit einer *abzählbaren* Teilmenge  $M \subset \mathbb{R}$ , nennen wir eine diskrete reelle Zufallsvariable.

Die Verteilung einer solchen Zufallsvariablen  $X$  ist vollständig durch die Wahrscheinlichkeiten

$$P(X = x) \quad \text{für alle } x \in M$$

beschrieben. Im gegenwärtigen Kontext sehen wir aber  $X$  als eine (spezielle) reelle Zufallsvariable an und ihre Verteilung daher als eine W-Verteilung auf ganz  $\mathbb{R}$  an; wir blähen also den Wertevorrat der Zufallsvariablen  $X$  von  $M$  zu ganz  $\mathbb{R}$  auf. Damit schaffen wir eine Situation wie in Bemerkung 1.1 von Kap. 1 angesprochen: Die Menge  $\bar{M} = \mathbb{R} \setminus M$  ist redundant (bezogen auf die Zufallsvariable  $X$ ), da das Ereignis  $\{X \in \bar{M}\}$  die leere Menge ist, und folglich  $P(X \in \bar{M}) = 0$ . Diese Redundanz nehmen wir in Kauf, um eine (soweit möglich) einheitliche Behandlung diskreter und nicht-diskreter reeller Zufallsvariablen zu haben.

Die Verteilungsfunktion  $F$  (gemäß der allgemeinen Definition 3.1) einer diskreten reellen Zufallsvariablen  $X$  ist eine monoton wachsende Treppenfunktion, die ihre Sprungstellen in den Werten  $x \in M$  hat und die zwischen je zwei benachbarten Werten aus  $M$  konstant ist. Wir wollen dies genauer erläutern im Fall, dass die Menge  $M \subset \mathbb{R}$  endlich ist. Sei also

$$X : \Omega \longrightarrow M = \{x_1, x_2, \dots, x_m\} ,$$

wobei  $x_1, x_2, \dots, x_m$  verschiedene reelle Zahlen sind, die wir hier als aufsteigend geordnet annehmen, d.h.

$$x_1 < x_2 < \dots < x_m .$$

Dann erhalten wir für ein beliebiges  $x \in \mathbb{R}$ :

$$F(x) = P(X \leq x) = P\left(\bigcup_{i: x_i \leq x} \{X = x_i\}\right) = \sum_{i: x_i \leq x} P(X = x_i) ,$$

und wir erhalten für die *Verteilungsfunktion*  $F$  einer *diskreten reellen Zufallsvariablen*:

$$F(x) = \begin{cases} 0 & , \text{ falls } x < x_1 \\ \sum_{i=1}^k P(X = x_i) & , \text{ falls } x_k \leq x < x_{k+1} , k \in \{1, \dots, m-1\} \\ 1 & , \text{ falls } x \geq x_m . \end{cases} \quad (3.2)$$

Für eine binomial- $(n = 6, p = 0.4)$ -verteilte Zufallsvariable  $X$  ist die Verteilungsfunktion  $F$  in Abbildung 3.2 dargestellt (hier ist  $m = 7, M = \{0, 1, \dots, 6\}$ ).

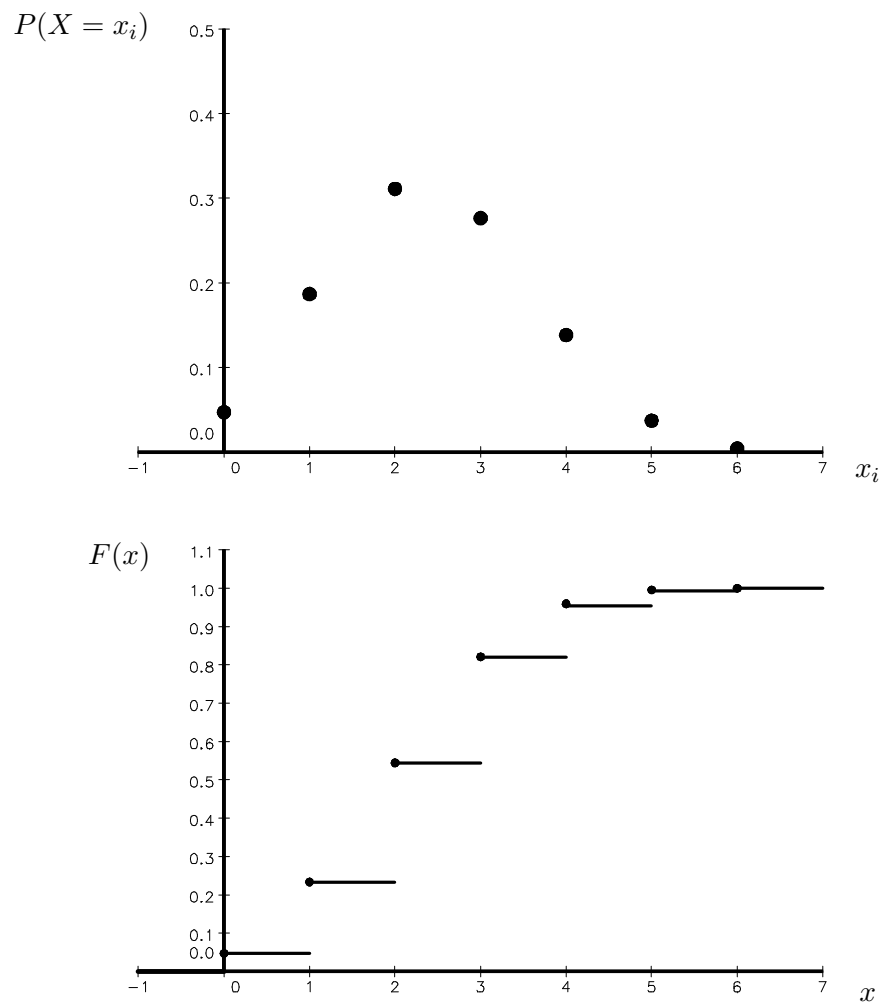


Abb. 3.2: Einzelwahrscheinlichkeiten (oben) und Verteilungsfunktion (unten) einer binomial- $(6, 0.4)$ -verteilten Zufallsvariablen

### 3.3 Stetig-verteilte Zufallsvariablen

Eine reelle Zufallsvariable  $X : \Omega \longrightarrow \mathbb{R}$  heißt stetig-verteilt, wenn ihre Verteilungsfunktion  $F$  stetig ist. Nach Bemerkung 3.3 ist dies äquivalent mit

$$P(X = x) = 0 \quad \text{für alle } x \in \mathbb{R}.$$

Daraus folgt auch, dass für eine stetig-verteilte Zufallsvariable  $X$  gilt:

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = F(b) - F(a)$$

für alle  $a, b \in \mathbb{R}$  mit  $a < b$ , sowie

$$\begin{aligned} P(X \leq b) &= P(X < b) = F(b) \quad \text{für alle } b \in \mathbb{R}, \\ P(X \geq a) &= P(X > a) = 1 - F(a) \quad \text{für alle } a \in \mathbb{R}. \end{aligned}$$

Zum Beispiel ist ja

$$P(a \leq X \leq b) = P(a < X \leq b) + \underbrace{P(X = a)}_{=0} = F(b) - F(a),$$

wobei die letzte Gleichung aus Theorem 3.2. Punkt (v), herrührt.

In den wichtigsten Fällen stetig-verteilter reeller Zufallsvariablen ist die Verteilungsfunktion nicht nur stetig, sondern sogar differenzierbar (zumindest an “fast allen” Stellen  $x \in \mathbb{R}$ , d.h. an allen bis auf höchstens endlich vielen Stellen  $x \in \mathbb{R}$ ). Dann haben wir die Ableitung

$$f(x) = F'(x) \quad \text{für fast alle } x \in \mathbb{R}.$$

Definieren wir noch die Werte von  $f$  an den evtl. vorhandenen (endlich vielen) Nicht-Differenzierbarkeits-Stellen von  $F$  irgendwie, aber  $\geq 0$  (z.B.  $f(x) = 0$  für diese  $x$ ), dann haben wir eine Funktion  $f : \mathbb{R} \longrightarrow \mathbb{R}$ , die nicht-negativ ist ( $f(x) \geq 0$  für alle  $x \in \mathbb{R}$ ), da ja die Verteilungsfunktion  $F$  monoton wachsend ist (Theorem 3.2, Punkt (i)) und ihre Ableitung daher nicht-negativ ist.

#### Definition 3.4 (Absolut-stetig-verteilte ZV’e, Dichtefunktion)

Wenn  $X$  eine stetig-verteilte reelle Zufallsvariable ist, deren Verteilungsfunktion  $F$  (fast überall) differenzierbar ist, dann heißt  $X$  eine absolut-stetig-verteilte Zufallsvariable. Die Ableitung  $f$  von  $F$  (evtl. im etwas erweiterten Sinne, s. oben) heißt dann die Dichtefunktion der Zufallsvariablen  $X$ .

#### Theorem 3.5 (Eigenschaften der Dichtefunktion)

Sei  $X$  eine absolut-stetig-verteilte reelle Zufallsvariable. Dann gilt für ihre Dichtefunktion  $f$  und ihre Verteilungsfunktion  $F$ :

$$\begin{aligned} f(x) &\geq 0 \quad \text{für alle } x \in \mathbb{R}, \\ \int_{-\infty}^{\infty} f(x) \, dx &= 1, \\ \int_{-\infty}^z f(x) \, dx &= F(z) \quad \text{für alle } z \in \mathbb{R}, \\ \int_a^b f(x) \, dx &= F(b) - F(a) \quad \text{für alle } a, b \in \mathbb{R} \text{ mit } a < b. \end{aligned}$$

In diesem Fall lassen sich also Wahrscheinlichkeiten

$$P(X \leq z) = F(z) \quad (\text{für } z \in \mathbb{R}), \text{ und} \\ P(a \leq X \leq b) = F(b) - F(a) \quad (\text{für } a, b \in \mathbb{R}), a < b, \quad \text{usw.}$$

als (bestimmte, evtl. uneigentliche) Integrale der Dichtefunktion und daher anschaulich als Flächeninhalte (s. Abb. 3.3) beschreiben.

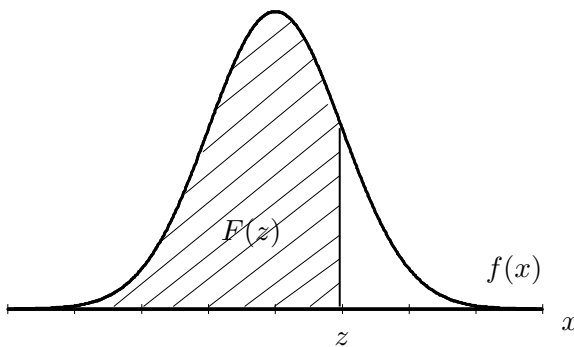


Abb. 3.3 (a)

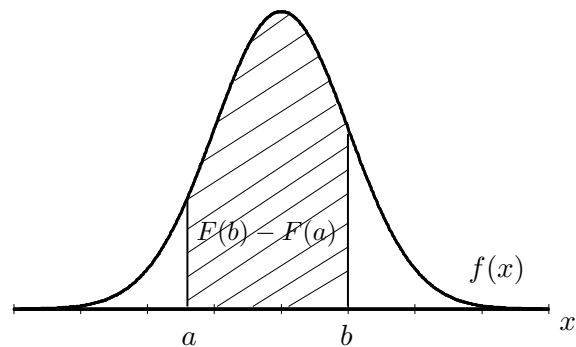


Abb. 3.3 (b)

**Beachte aber:** Die Werte  $f(x)$  der Dichtefunktion selbst stellen keine Wahrscheinlichkeiten dar; Werte der Dichtefunktion können insbesondere auch größer als 1 sein.

Die folgende Definition präsentiert drei der wichtigsten Klassen von absolut-stetig-verteilten Zufallsvariablen.

**Definition 3.6 (Wichtige Klassen abs.-stetig-verteilter ZV'en)**

Sei  $X$  eine absolut-stetig-verteilte reelle Zufallsvariable.

- (a) Seien  $\mu \in \mathbb{R}$  und  $\sigma > 0$ . Die Zufallsvariable  $X$  heißt normal- $(\mu, \sigma)$ -verteilt, oder auch Gauß- $(\mu, \sigma)$ -verteilt, wenn ihre Dichtefunktion gegeben ist durch

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \quad \text{für alle } x \in \mathbb{R}.$$

(Hier bezeichnet  $\pi$  die Kreiszahl,  $\pi \approx 3.14159$ ).

- (b) Seien  $\alpha, \beta \in \mathbb{R}$  mit  $\alpha < \beta$ . Die Zufallsvariable  $X$  heißt gleichverteilt auf dem Intervall  $[\alpha, \beta]$ , oder auch Rechteck- $(\alpha, \beta)$ -verteilt, wenn ihre Dichtefunktion gegeben ist durch

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & , \text{ falls } \alpha \leq x \leq \beta \\ 0 & , \text{ sonst} \end{cases} \quad (x \in \mathbb{R}).$$

- (c) Sei  $\lambda > 0$ . Die Zufallsvariable  $X$  heißt exponential- $(\lambda)$ -verteilt, wenn ihre Dichtefunktion gegeben ist durch

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & , \text{ falls } x > 0 \\ 0 & , \text{ sonst} \end{cases} \quad (x \in \mathbb{R}).$$

Die Verteilungsfunktionen für die in Definition 3.6 genannten Fälle ergeben sich mit Hilfe von Theorem 3.5 wie folgt:

(a) Normal- $(\mu, \sigma)$ -verteilte Zufallsvariable:

Zunächst der Fall  $\mu = 0$  und  $\sigma = 1$ ; man spricht dann von einer standard-normal-verteilten Zufallsvariablen, und ihre Verteilungsfunktion bezeichnet man üblicherweise mit dem griechischen Buchstaben  $\Phi$ . Mit Theorem 3.5 ist

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{1}{2}x^2\right) dx ,$$

und diese Formel lässt sich nicht “expliziter” schreiben. Die Funktion  $\Phi$  liegt jedoch in tabellierter (s. Anhang, Tafel A.1) und implementierter Form vor.

Für beliebige Parameterwerte  $\mu \in \mathbb{R}$  und  $\sigma > 0$  können wir die Verteilungsfunktion  $F$  einer normal- $(\mu, \sigma)$ -verteilten Zufallsvariablen durch  $\Phi$  ausdrücken gemäß

$$F(z) = \Phi\left(\frac{z - \mu}{\sigma}\right) \quad \text{für alle } z \in \mathbb{R}. \quad (3.3)$$

Diesen Zusammenhang zwischen der Verteilungsfunktion einer normal- $(\mu, \sigma)$ -verteilten und einer standard-normalverteilten Zufallsvariablen erhält man durch Anwendung der Substitutionsregel für Integrale in

$$F(z) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right) dx ,$$

mit der Substitution  $y = (x - \mu)/\sigma$ .

(b) Gleichverteilte Zufallsvariable auf  $[\alpha, \beta]$ :

$$F(z) = \begin{cases} 0 & , \text{ falls } z < \alpha \\ \frac{z - \alpha}{\beta - \alpha} & , \text{ falls } \alpha \leq z \leq \beta \\ 1 & , \text{ falls } z > \beta \end{cases}$$

Schreiben wir nämlich die Dichtefunktion aus Definition 3.6 (b) in kompakterer Form als

$$f(x) = \frac{1}{\beta - \alpha} \mathbf{1}_{[\alpha, \beta]}(x) , \quad \text{für alle } x \in \mathbb{R} ,$$

mit der sog. *Indikatorfunktion* des Intervalles  $[\alpha, \beta]$ , definiert durch  $\mathbf{1}_{[\alpha, \beta]}(x) = \begin{cases} 1, & \text{falls } x \in [\alpha, \beta] \\ 0, & \text{falls } x \notin [\alpha, \beta] \end{cases}$ , dann ist

$$F(z) = \frac{1}{\beta - \alpha} \int_{-\infty}^z \mathbf{1}_{[\alpha, \beta]}(x) dx \quad \text{für alle } z \in \mathbb{R} ;$$

die Auswertung des Integrals (mit den nötigen Fallunterscheidungen für  $z$ ) ergibt die behauptete Formel.

(c) Exponential- $(\lambda)$ -verteilte Zufallsvariable:

$$F(z) = \begin{cases} 1 - \exp(-\lambda z) & , \text{ falls } z > 0 \\ 0 & , \text{ sonst} \end{cases} \quad (z \in \mathbb{R}). \quad (3.4)$$

Eine kompaktere Darstellung der Dichtefunktion aus Definition 3.6 (c) ergibt nämlich,

$$F(z) = \lambda \int_{-\infty}^z \exp(-\lambda x) \mathbf{1}_{]0, \infty[}(x) dx ,$$

wobei wir jetzt die Indikatorfunktion des Intervalles  $]0, \infty[$  verwenden. Wenn  $z \leq 0$ , dann ist der Integrand auf dem Integrationsbereich (von  $-\infty$  bis  $z$ ) offensichtlich konstant gleich Null, folglich ist  $F(z) = 0$ . Wenn  $z > 0$ , dann liefert die Integration von  $-\infty$  bis 0 aus demselben Grund den Wert Null; übrig bleibt die Integration von 0 bis  $z$ , also ist

$$F(z) = \lambda \int_0^z \exp(-\lambda x) dx = \left[ -\exp(-\lambda x) \right]_{x=0}^{x=z} = 1 - \exp(-\lambda z) .$$

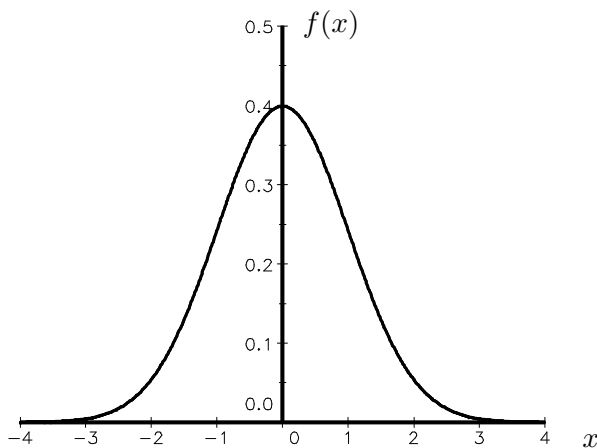
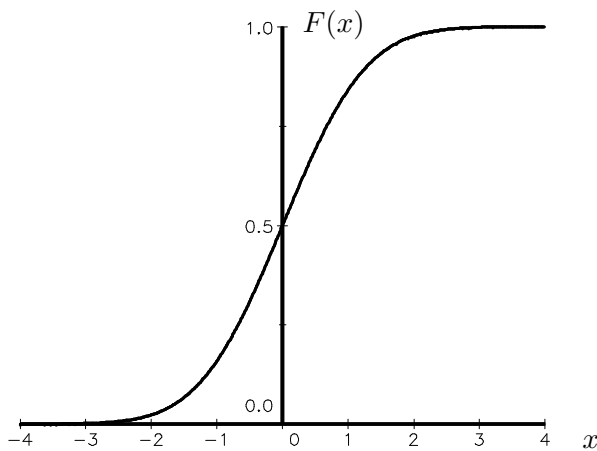


Abb. 3.4: Verteilungsfunktion (oben) und Dichtefunktion (unten) einer normalverteilten Zufallsvariablen ( $\mu = 0, \sigma = 1$ )

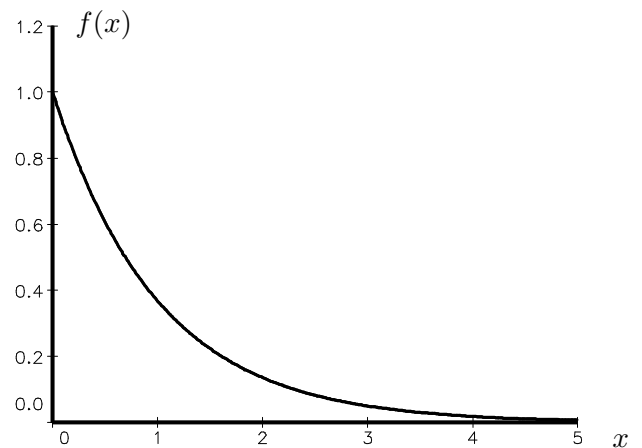
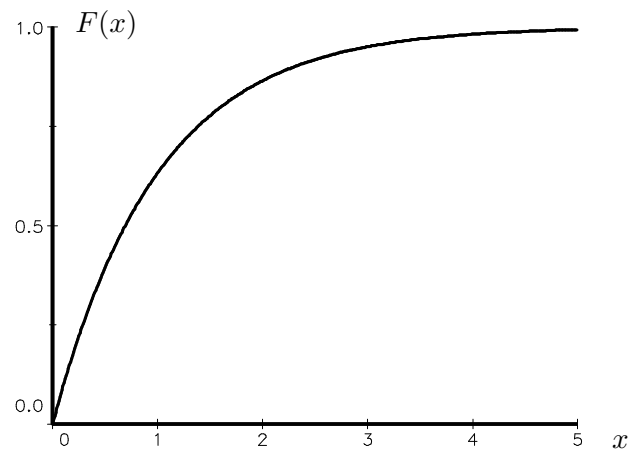


Abb. 3.5: Verteilungsfunktion (oben) und Dichtefunktion (unten) einer exponentialverteilten Zufallsvariablen ( $\lambda = 1$ )

### Bemerkung 3.7 (Lineare Transformationen)

Betrachten wir zu einer reellen Zufallsvariablen  $X$  und zu reellen Zahlen  $a \neq 0$  und  $c$  die *linear transformierte* (neue) Zufallsvariable

$$Y = aX + c , \quad \text{d.h.} \quad Y(\omega) = aX(\omega) + c \quad \text{für alle } \omega \in \Omega .$$

Wenn  $X$  eine normal- $(\mu, \sigma)$ -verteilte Zufallsvariable ist, dann ist die linear transformierte Zufallsvariable  $Y = aX + c$  ebenfalls normal-verteilt, und zwar zu den Parameterwerten  $\mu_Y = a\mu + c$  und  $\sigma_Y = |a|\sigma$ .

Das können wir so einsehen (wir beschränken uns auf den Fall  $a > 0$ ):

Bezeichne  $F_Y$  die Verteilungsfunktion der Zufallsvariablen  $Y$  und  $F_X$  die Verteilungsfunktion der Zufallsvariablen  $X$ . Für jedes  $z \in \mathbb{R}$  ist

$$\begin{aligned} F_Y(z) &= P(Y \leq z) = P(aX + c \leq z) = P\left(X \leq \frac{z-c}{a}\right) \\ &= F_X\left(\frac{z-c}{a}\right) = \Phi\left(\frac{\frac{z-c}{a} - \mu}{\sigma}\right) = \Phi\left(\frac{z - (a\mu + c)}{a\sigma}\right), \end{aligned}$$

und dies ist ja die Verteilungsfunktion einer normal- $(a\mu + c, a\sigma)$ -verteilten Zufallsvariablen.

Von besonderer Bedeutung ist diese Eigenschaft im Fall  $a = 1/\sigma$ ,  $c = -\mu/\sigma$ . Dann ergibt sich:

Die Zufallsvariable  $Y = \frac{X - \mu}{\sigma}$  ist standard-normalverteilt, wenn  $X$  normal- $(\mu, \sigma)$ -verteilt ist.

Ähnliches lässt sich auch im Falle einer (auf einem Intervall  $[\alpha, \beta]$ ) gleichverteilten Zufallsvariablen  $X$  sehen: Die linear transformierte Zufallsvariable  $Y$  ist wiederum gleichverteilt, jetzt auf dem entsprechend linear transformierten Intervall.

Für eine exponential- $(\lambda)$ -verteilte Zufallsvariable  $X$  gilt ein entsprechendes Resultat nur im Fall  $c = 0$  und  $a > 0$ : Dann ist die Zufallsvariable  $Y = aX$  exponential- $(\lambda/a)$ -verteilt.

### Bemerkung 3.8 (Anwendung in der Modellierung)

Die in Definition 3.6 genannten Verteilungen werden oft zur Modellierung von Zufallsexperimenten mit kontinuierlichen Ergebnissen verwendet.

Normalverteilte Zufallsvariablen haben einen fast universell erscheinenden Anwendungsbereich (z.B. eine fehlerbehaftete Messung, eine Produkt-Charakteristik in der industriellen Fertigung, eine Projektdauer, eine Nachfrage nach einem gehandelten Produkt).

Gleichverteilte (auf einem Intervall  $[\alpha, \beta]$ ) Zufallsvariablen bilden ein stetiges Analogon zu einer diskreten Gleichverteilung. Das Ergebnis des Zufallsexperiments wird als eine "völlig zufällig" gezogene Zahl aus dem Intervall  $[\alpha, \beta]$  interpretiert. Anwendung z.B. bei der Modellierung einer zufälligen Ankunfts- oder Wartezeit, oder auch einer fehlerbehafteten Messung; die Standard-Zufallszahlen, die mit einem Computer erzeugt werden, sind Werte von gleichverteilten Zufallsvariablen auf  $[0, 1]$ .

Exponential-verteilte Zufallsvariablen werden insbesondere zur Modellierung zufälliger kontinuierlicher Zeitdauern verwendet (Lebensdauern, Projektdauern, Wartezeiten).



# Kapitel 4

## Erwartungswert, Varianz und Standardabweichung

### 4.1 Erwartungswert

Der Erwartungswert einer reellen Zufallsvariablen  $X$  ist eine wichtige Kenngröße der Verteilung von  $X$ . Er ist ein Mittelwert aller möglichen Werte, welche die Zufallsvariable annehmen kann. Man nennt den Erwartungswert auch einen Lageparameter der Verteilung der Zufallsvariablen, da er eine Kenngröße für die Lage (Position) der Verteilung von  $X$  auf der reellen Zahlengeraden ist. Eine andere wichtige Kenngröße der Verteilung von  $X$  ist die Varianz bzw. die Standardabweichung (s. Abschnitt 4.2 unten); die Standardabweichung (und auch die Varianz) ist eine Kenngröße für die Streubreite der Verteilung um ihren Erwartungswert, und man spricht daher auch von einem Streuungsparameter der Verteilung von  $X$ .

Um zu einer präzisen Definition des Erwartungswertes zu kommen, betrachten wir als erstes den Fall einer *diskreten* reellen Zufallsvariablen:

$$X : \Omega \longrightarrow M \subset \mathbb{R} ,$$

wobei zunächst  $M$  eine *endliche* Menge bestehend aus  $m$  Zahlen sei,  $M = \{x_1, \dots, x_m\}$ . Die möglichen Werte der Zufallsvariablen  $X$  sind also  $x_1, \dots, x_m$ . Der Erwartungswert von  $X$  ist nun der Mittelwert dieser Werte, allerdings (im Allgemeinen) nicht das gewöhnliche arithmetische Mittel  $\frac{1}{m} \sum_{i=1}^m x_i$ , sondern das entsprechend den Wahrscheinlichkeiten gewichtete Mittel,

$$E(X) = \sum_{i=1}^m x_i P(X = x_i) .$$

#### Beispiel 4.1

(a) Sei  $X$  eine auf  $M$  gleichverteilte Zufallsvariable, d.h.

$$P(X = x_i) = 1/m \quad \text{für alle } i = 1, \dots, m ,$$

In diesem speziellen Fall erhalten wir als Erwartungswert tatsächlich das *arithmetische* Mittel:

$$E(X) = \frac{1}{m} \sum_{i=1}^m x_i = \bar{x} .$$

(b) Sei  $X$  eine 0-1-wertige Zufallsvariable, also  $M = \{0, 1\}$ , folglich

$$E(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = P(X = 1) .$$

Für eine 0-1-Variable  $X : \Omega \longrightarrow \{0, 1\}$  gilt also

$$E(X) = P(X = 1) . \quad \blacksquare$$

Als nächstes betrachten wir eine diskrete reelle Zufallsvariable, die abzählbar-unendlich viele verschiedenen Werte annehmen kann,

$$X : \Omega \longrightarrow M \subset \mathbb{R} ,$$

wobei  $M$  aus den unendlich vielen, aber nummerierbaren Zahlen  $x_1, x_2, \dots, x_i, \dots$  (alle verschieden) besteht (z.B.  $M = \mathbb{N}$ ). Dann ist der Erwartungswert von  $X$  das entsprechend den Wahrscheinlichkeiten gewichtete Mittel der Werte  $x_i$ ,

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i) ,$$

(dabei setzen wir stillschweigend die Konvergenz dieser unendlichen Reihe voraus).

Unsere bisherigen Ausführungen können wir wie folgt zusammenfassen:

**Definition 4.2 (Erwartungswert einer diskreten Zufallsvariablen)**

Sei  $X$  eine diskrete reelle Zufallsvariable, d.h.  $X : \Omega \longrightarrow M$ , wobei  $M$  eine abzählbare Teilmenge von  $\mathbb{R}$  ist. Dann heißt die reelle Zahl

$$E(X) = \sum_{x \in M} x P(X = x)$$

der Erwartungswert der Zufallsvariablen  $X$ .

**Beispiel 4.3 (Diskrete Zufallsvariablen aus Definition 1.15)**

- (a) Eine auf einer endlichen Teilmenge  $M \subset \mathbb{R}$  gleichverteilte Zufallsvariable  $X$ :  
Dies wurde schon in Beispiel 4.1(a) betrachtet; es ergab sich  $E(X) = \bar{x}$ , das arithmetische Mittel der Werte in  $M$ .
- (b) Eine binomial-( $n, p$ )-verteilte Zufallsvariable  $X$ :  
Hier ist  $M = \{0, 1, \dots, n\}$  und

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \quad \text{für } i = 0, 1, \dots, n ;$$

folglich

$$E(X) = \sum_{i=0}^n i \binom{n}{i} p^i (1-p)^{n-i} .$$

Für  $i \geq 1$  schreiben wir

$$i \binom{n}{i} = i \frac{n!}{i! (n-i)!} = n \frac{(n-1)!}{(i-1)! ((n-1)-(i-1))!} = n \binom{n-1}{i-1} ,$$

$$p^i (1-p)^{n-i} = p p^{i-1} (1-p)^{(n-1)-(i-1)} ,$$

und wir erhalten

$$\begin{aligned} E(X) &= np \sum_{i=1}^n \binom{n-1}{i-1} p^{i-1} (1-p)^{(n-1)-(i-1)} \\ &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{(n-1)-j} = np, \end{aligned}$$

da die Summe in der letzten Zeile nach der binomischen Formel gleich

$$(p + (1-p))^{n-1} = 1$$

ist.

- (c) Eine hypergeometrisch- $(N, s, n)$ -verteilte Zufallsvariable  $X$ :  
Hier ist  $M = \{0, 1, \dots, n\}$  und

$$P(X = i) = \frac{\binom{s}{i} \binom{N-s}{n-i}}{\binom{N}{n}} \quad \text{für } i = 0, 1, \dots, n,$$

folglich

$$E(X) = \sum_{i=0}^n i \binom{s}{i} \binom{N-s}{n-i} / \binom{N}{n}.$$

Mit geeigneten Umformungen (auf deren Darstellung wir hier verzichten) ergibt sich

$$E(X) = \frac{ns}{N}.$$

- (d) Eine Poisson- $(\lambda)$ -verteilte Zufallsvariable  $X$ :  
Hier ist  $M = \mathbb{N}_0$  und

$$P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!} \quad \text{für alle } i \in \mathbb{N}_0,$$

folglich

$$E(X) = \sum_{i=0}^{\infty} i e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \lambda \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = e^{-\lambda} \lambda \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}}_{=e^\lambda} = \lambda.$$

- (e) Eine geometrisch- $(p)$ -verteilte Zufallsvariable  $X$ :  
Hier ist  $M = \mathbb{N}$  und

$$P(X = i) = p(1-p)^{i-1} \quad \text{für alle } i \in \mathbb{N},$$

folglich

$$E(X) = \sum_{i=1}^{\infty} i p (1-p)^{i-1} = p \sum_{i=1}^{\infty} i (1-p)^{i-1}.$$

Es lässt sich mit Hilfe von Resultaten über Potenzreihen zeigen, dass

$$\sum_{i=1}^{\infty} i (1-p)^{i-1} = \frac{1}{p^2},$$

und wir erhalten damit

$$E(X) = \frac{1}{p}.$$

■

Kommen wir nun zum allgemeinen Fall einer *beliebigen* reellen Zufallsvariablen  $X : \Omega \longrightarrow \mathbb{R}$ , die also auch nicht-diskret (insbesondere stetig-verteilt) sein kann. Ein theoretisches Resultat in diesem Zusammenhang besagt, dass die Zufallsvariable  $X$  durch eine Folge diskreter reeller Zufallsvariablen approximiert werden kann:

**Theorem 4.4**

Sei  $X : \Omega \longrightarrow \mathbb{R}$  eine reelle Zufallsvariable. Dann existiert eine Folge diskreter reeller Zufallsvariablen  $X_n : \Omega \longrightarrow M_n \subset \mathbb{R}$ , ( $n = 1, 2, \dots$ ), die gegen die Zufallsvariable  $X$  im folgenden Sinne konvergiert:

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \quad \text{für alle } \omega \in \Omega.$$

**Definition 4.5 (Erwartungswert einer reellen Zufallsvariablen)**

Sei  $X : \Omega \longrightarrow \mathbb{R}$  eine reelle Zufallsvariable. Sei  $X_n$  ( $n = 1, 2, \dots$ ) eine Folge diskreter Zufallsvariablen auf  $\Omega$  gemäß Theorem 4.4. Dann heißt

$$E(X) = \lim_{n \rightarrow \infty} E(X_n)$$

der Erwartungswert der Zufallsvariablen  $X$ .

Dies ist natürlich eine theoretische Definition, die sich zur Berechnung von Erwartungswerten in konkreten Fällen wenig eignet. Die für uns wichtigste Klasse nicht-diskreter reeller Zufallsvariablen bilden die absolut-stetig-verteilten Zufallsvariablen. Hier lässt sich eine Integralformel für den Erwartungswert mit Hilfe der Dichtefunktion formulieren.

**Theorem 4.6 (Erwartungswert einer absolut-stetig-verteilten Zufallsvariablen)**

Sei  $X : \Omega \longrightarrow \mathbb{R}$  eine absolut-stetig-verteilte Zufallsvariable mit der Dichtefunktion  $f$ . Dann gilt:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

**Beispiel 4.7 (Absolut-stetig-verteilte Zufallsvariablen aus Def. 3.6)**

(a) Eine normal- $(\mu, \sigma)$ -verteilte Zufallsvariable  $X$ :

Wir behandeln zuerst den Fall  $\mu = 0$ ,  $\sigma = 1$  (standard-normal-verteilte Zufallsvariable). Nach Theorem 4.6 haben wir

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2}x^2\right) dx.$$

Die Funktion  $-\exp\left(-\frac{1}{2}x^2\right)$  ist eine Stammfunktion des Integranden  $x \exp\left(-\frac{1}{2}x^2\right)$ , folglich

$$\int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2}x^2\right) dx = \left[-\exp\left(-\frac{1}{2}x^2\right)\right]_{x=-\infty}^{x=\infty} = -0 + 0 = 0.$$

Der Erwartungswert einer standard-normal-verteilten Zufallsvariablen ist also gleich Null. Seien jetzt  $\mu \in \mathbb{R}$  und  $\sigma > 0$  beliebig. Offensichtlich besteht zwischen der Dichtefunktion

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

einer normal- $(\mu, \sigma)$ -verteilten Zufallsvariablen  $X$  und der Dichtefunktion

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

einer standard-normal-verteilten Zufallsvariablen  $X_0$  der folgende Zusammenhang:

$$\boxed{f(x) = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right)} \quad (4.1)$$

Wir erhalten, unter Benutzung der Substitutionsregel für Integrale, (mit der Substitution  $y = (x - \mu)/\sigma$ ),

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) dx = \int_{-\infty}^{\infty} (\sigma y + \mu) \phi(y) dy \\ &= \sigma \int_{-\infty}^{\infty} y \phi(y) dy + \mu \int_{-\infty}^{\infty} \phi(y) dy . \end{aligned}$$

Nun ist nach dem oben Gezeigten

$$\int_{-\infty}^{\infty} y \phi(y) dy = E(X_0) = 0 ,$$

und nach Theorem 3.5 haben wir

$$\int_{-\infty}^{\infty} \phi(y) dy = 1 .$$

Damit erhalten wir  $E(X) = \mu$ , der Erwartungswert einer normal- $(\mu, \sigma)$ -verteilten Zufallsvariablen ist also gleich  $\mu$ .

- (b) Eine auf  $[\alpha, \beta]$  gleichverteilte Zufallsvariable  $X$ :  
Die Dichtefunktion ist

$$f(x) = \frac{1}{\beta - \alpha} \mathbf{1}_{[\alpha, \beta]}(x) ,$$

folglich nach Theorem 4.6:

$$\begin{aligned} E(X) &= \frac{1}{\beta - \alpha} \int_{-\infty}^{\infty} x \mathbf{1}_{[\alpha, \beta]}(x) dx = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} x dx \\ &= \frac{1}{\beta - \alpha} \left( \frac{\beta^2}{2} - \frac{\alpha^2}{2} \right) = \frac{1}{2} \frac{(\beta + \alpha)(\beta - \alpha)}{\beta - \alpha} = \frac{1}{2} (\beta + \alpha) . \end{aligned}$$

Der Erwartungswert einer auf dem Intervall  $[\alpha, \beta]$  gleichverteilten Zufallsvariablen ist also der Mittelpunkt  $(\alpha + \beta)/2$  des Intervalls.

- (c) Eine exponential- $(\lambda)$ -verteilte Zufallsvariable  $X$ :  
Die Dichtefunktion ist gleich

$$f(x) = \lambda \exp(-\lambda x) \mathbf{1}_{]0, \infty[}(x) ,$$

folglich nach Theorem 4.6:

$$E(X) = \int_{-\infty}^{\infty} x \lambda \exp(-\lambda x) \mathbf{1}_{]0, \infty[}(x) dx = \int_0^{\infty} x \cdot (\lambda \exp(-\lambda x)) dx .$$

Mit partieller Integration erhält man als eine Stammfunktion des Integranden die Funktion

$$-x \exp(-\lambda x) - \frac{1}{\lambda} \exp(-\lambda x) ,$$

und folglich

$$E(X) = \left[ -x \exp(-\lambda x) - \frac{1}{\lambda} \exp(-\lambda x) \right]_{x=0}^{x=\infty} = -0 - 0 + 0 + \frac{1}{\lambda} .$$

Der Erwartungswert einer exponential-( $\lambda$ )-verteilten Zufallsvariablen ist also gleich  $1/\lambda$ . ■

Wir präsentieren nun allgemeine Rechenregeln für Erwartungswerte (Linearität, Monotonie) sowie Formeln für Erwartungswerte von transformierten Zufallsvariablen.

Der Erwartungswert verhält sich linear bezüglich Linearkombinationen mehrerer reeller Zufallsvariablen auf  $\Omega$ . Betrachten wir zunächst zwei reelle Zufallsvariablen

$$X : \Omega \longrightarrow \mathbb{R} \quad \text{und} \quad Y : \Omega \longrightarrow \mathbb{R} .$$

Seien noch  $a$ ,  $b$  und  $c$  reelle Zahlen. Dann ist durch

$$aX(\omega) + bY(\omega) + c \quad \text{für alle } \omega \in \Omega$$

wiederum eine reelle Zufallsvariable auf  $\Omega$  gegeben, die Linearkombination  $aX + bY + c$ . Entsprechend können wir für mehrere reelle Zufallsvariablen  $X_1, X_2, \dots, X_n$  auf  $\Omega$  und reelle Zahlen  $a_1, a_2, \dots, a_n$  und  $c$  die Linearkombination

$$\sum_{i=1}^n a_i X_i + c$$

bilden. Es gilt das folgende Resultat:

#### Theorem 4.8 (Linearität des Erwartungswertes)

$$E(aX + bY + c) = aE(X) + bE(Y) + c ,$$

insbesondere

$$E(aX + c) = aE(X) + c ;$$

allgemeiner gilt

$$E\left(\sum_{i=1}^n a_i X_i + c\right) = \sum_{i=1}^n a_i E(X_i) + c .$$

#### Theorem 4.9 (Monotonie des Erwartungswertes)

Seien  $X$  und  $Y$  reelle Zufallsvariablen auf  $\Omega$ . Sei  $c \in \mathbb{R}$ .

- (i) Gilt  $X(\omega) \leq Y(\omega)$  für alle  $\omega \in \Omega$ , dann ist  $E(X) \leq E(Y)$ .
- (ii) Gilt  $X(\omega) \leq c$  für alle  $\omega \in \Omega$ , dann ist  $E(X) \leq c$ .
- (iii) Gilt  $X(\omega) \geq c$  für alle  $\omega \in \Omega$ , dann ist  $E(X) \geq c$ .

Unter einer Transformation einer reellen Zufallsvariablen  $X$  verstehen wir eine (neue) reelle Zufallsvariable  $Y$  der Form

$$Y = g(X) , \quad \text{d.h.} \quad Y(\omega) = g(X(\omega)) \quad \text{für alle } \omega \in \Omega ,$$

wobei  $g : \mathbb{R} \longrightarrow \mathbb{R}$  eine Funktion ("Transformation") ist.

Im Fall, dass  $g$  eine lineare Funktion ist, d.h.

$$g(x) = ax + c \quad \text{für alle } x \in \mathbb{R}$$

mit gewissen reellen Konstanten  $a$  und  $c$ , haben wir  $g(X) = aX + c$ , und wir erhalten mit der zweiten Formel von Theorem 4.8:

$$E(g(X)) = aE(X) + c = g(E(X)) ,$$

d.h. Erwartungswert und Transformation dürfen vertauscht werden. Das ist jedoch nur im Falle einer linearen Transformation richtig.

Wenn  $g$  eine nicht-lineare Transformation ist, z.B.  $g(x) = x^2$ , dann ist im Allgemeinen

$$E(g(X)) \neq g(E(X)) ;$$

bestenfalls kann die rechte Seite als eine (meist grobe) Approximation für den Erwartungswert der transformierten Zufallsvariablen  $Y = g(X)$  angesehen werden. Zum Beispiel gilt für die quadratische Funktion  $g(x) = x^2$  stets

$$E(X^2) \geq (E(X))^2 ,$$

und oft ist die linke Seite sogar *erheblich* größer als die rechte Seite (s. auch Abschnitt 4.2 im Zusammenhang mit Varianzen).

Zur (exakten) Berechnung des Erwartungswertes einer nicht-linear transformierten Zufallsvariablen ist das folgende Resultat nützlich.

#### Theorem 4.10 (Erwartungswert transformierter Zufallsvariablen)

Seien  $g : \mathbb{R} \longrightarrow \mathbb{R}$  eine Funktion und  $X : \Omega \longrightarrow \mathbb{R}$  eine reelle Zufallsvariable auf  $\Omega$ .

- (a) Wenn die Zufallsvariable  $X$  diskret ist, d.h.  $X : \Omega \longrightarrow M \subset \mathbb{R}$  mit einer abzählbaren Menge  $M$ , dann gilt

$$E(g(X)) = \sum_{x \in M} g(x) P(X = x) .$$

- (b) Wenn  $X$  eine absolut-stetig-verteilte Zufallsvariable mit der Dichtefunktion  $f$  ist, dann gilt

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx .$$

#### Beispiel 4.11 (Zweites Moment einer auf einem Intervall gleichverteilten ZV'en)

Sei  $X$  eine auf dem Intervall  $[\alpha, \beta]$  gleichverteilte Zufallsvariable (wobei  $\alpha, \beta \in \mathbb{R}$  mit  $\alpha < \beta$  gegeben sind). Wir wollen  $E(X^2)$  (das sog. *zweite Moment* der Zufallsvariablen  $X$ ) berechnen. Anwendung von Theorem 4.10 (b) für die Transformation  $g(x) = x^2$  und die Dichtefunktion  $f(x) = \frac{1}{\beta - \alpha} \mathbf{1}_{[\alpha, \beta]}(x)$  von  $X$  ergibt:

$$\begin{aligned} E(X^2) &= \frac{1}{\beta - \alpha} \int_{-\infty}^{\infty} x^2 \mathbf{1}_{[\alpha, \beta]}(x) dx = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} x^2 dx = \frac{1}{\beta - \alpha} \left[ \frac{1}{3} x^3 \right]_{x=\alpha}^{x=\beta} \\ &= \frac{1}{3} \frac{\beta^3 - \alpha^3}{\beta - \alpha} = \frac{1}{3} \frac{(\beta - \alpha)(\beta^2 + \alpha\beta + \alpha^2)}{\beta - \alpha} = \frac{1}{3} (\beta^2 + \alpha\beta + \alpha^2) . \end{aligned}$$

■

## 4.2 Standardabweichung und Varianz

Die Standardabweichung einer reellen Zufallsvariablen  $X$  ist eine weitere wichtige Kenngröße der Verteilung von  $X$ . Sie ist ein Streuungsparameter der Verteilung von  $X$  und beschreibt die Streubreite der Zufallsvariablen um ihren Erwartungswert  $E(X)$  herum. Die Standardabweichung gibt eine mittlere Abweichung der möglichen Werte der Zufallsvariablen von ihrem Erwartungswert an. Dabei werden zunächst die *quadrierten* Abweichungen der Werte vom Erwartungswert genommen und daraus ein Mittelwert gebildet, der die Varianz der Zufallsvariablen genannt wird. Die Quadratwurzel daraus ist dann die Standardabweichung.

Mit dem bereits zur Verfügung stehenden Begriff des Erwartungswertes einer reellen Zufallsvariablen lässt sich dies wie folgt präzisieren.

Sei  $X : \Omega \longrightarrow \mathbb{R}$  eine reelle Zufallsvariable auf  $\Omega$ . Betrachte die transformierte Zufallsvariable

$$(X - E(X))^2,$$

d.h. die (neue) reelle Zufallsvariable, nennen wir sie für den Augenblick  $Y : \Omega \longrightarrow \mathbb{R}$ , die definiert ist durch

$$Y(\omega) = (X(\omega) - E(X))^2 \quad \text{für alle } \omega \in \Omega.$$

Die Werte dieser Zufallsvariablen  $Y$  sind offensichtlich die quadrierten Abweichungen der Werte der Zufallsvariablen  $X$  von ihrem Erwartungswert  $E(X)$ . Der Erwartungswert von  $Y$  ist dann die "mittlere quadrierte Abweichung" der (Werte der) Zufallsvariablen  $X$  von ihrem Erwartungswert.

### Definition 4.12 (Varianz/Standardabweichung einer reellen ZV'en)

Sei  $X : \Omega \longrightarrow \mathbb{R}$  eine reelle Zufallsvariable auf  $\Omega$ . Dann heißt die (nicht-negative) reelle Zahl

$$\text{Var}(X) = E\left((X - E(X))^2\right)$$

die Varianz der Zufallsvariablen  $X$ , und  $\sqrt{\text{Var}(X)}$  heißt die Standardabweichung der Zufallsvariablen  $X$ .

In der Tat ist stets  $\text{Var}(X) \geq 0$ , da  $(X - E(X))^2 \geq 0$  und folglich auch der Erwartungswert dieser Zufallsvariablen nicht-negativ (s. Theorem 4.9). Außer im Trivialfall, dass die Zufallsvariable  $X$  eine Konstante ist, gilt sogar stets

$$\text{Var}(X) > 0.$$

Eine zur Berechnung der Varianz oft nützliche Formel erhalten wir aus der Identität

$$(X(\omega) - E(X))^2 = (X(\omega))^2 - 2X(\omega)E(X) + (E(X))^2 \quad (\text{für alle } \omega \in \Omega),$$

oder kürzer geschrieben

$$(X - E(X))^2 = X^2 - 2E(X) \cdot X + (E(X))^2.$$

Die Linearität des Erwartungswertes (Theorem 4.8) ergibt daher (beachte:  $E(X)$  ist eine reelle Zahl),

$$\text{Var}(X) = E(X^2) - 2E(X) \cdot E(X) + (E(X))^2 = E(X^2) - (E(X))^2.$$

Damit haben wir auch die folgende Formel für die Varianz einer reellen Zufallsvariablen  $X$ :



**Theorem 4.13**

$$\text{Var}(X) = E(X^2) - (E(X))^2.$$

Der Erwartungswert  $E(X^2)$  heißt auch das zweite Moment der Zufallsvariablen  $X$  (während der Erwartungswert  $E(X)$  auch das erste Moment der Zufallsvariablen  $X$  genannt wird).

**Beispiel 4.14 (0-1-wertige Zufallsvariable)**

Sei  $X : \Omega \longrightarrow \{0, 1\}$  und bezeichne  $p = P(X = 1)$ . Offensichtlich ist  $X^2 = X$ , also  $E(X^2) = E(X) = p$  (s. Beispiel 4.1 (b)), und folglich nach Theorem 4.13

$$\text{Var}(X) = p - p^2 = p(1 - p).$$

Das Verhalten der Varianz bzw. der Standardabweichung bei einer *linearen Transformation* der Zufallsvariablen ist leicht zu ersehen:

Seien  $X : \Omega \longrightarrow \mathbb{R}$  eine reelle Zufallsvariable und  $a$  und  $c$  reelle Zahlen; wir betrachten die linear transformierte Zufallsvariable

$$Y = aX + c.$$

Wir wissen aus Theorem 4.8:

$$E(Y) = aE(X) + c.$$

Für die Varianz von  $Y$  erhalten wir aus Definition 4.12:

$$\begin{aligned} \text{Var}(Y) &= E\left([aX + c - (aE(X) + c)]^2\right) = E\left(a^2 [X - E(X)]^2\right) \\ &= a^2 E\left([X - E(X)]^2\right) = a^2 \text{Var}(X). \end{aligned}$$

Damit haben wir:

**Theorem 4.15**

$$\text{Var}(aX + c) = a^2 \text{Var}(X) \quad \text{bzw.} \quad \sqrt{\text{Var}(aX + c)} = |a| \sqrt{\text{Var}(X)}.$$

**Bemerkung 4.16 (Nichtlineare Transformationen)**

Bei einer nichtlinearen Transformation einer reellen Zufallsvariablen  $X$  zu  $Y = g(X)$  (mit einer Funktion  $g : \mathbb{R} \longrightarrow \mathbb{R}$ ) lässt sich die Varianz von  $Y$  in der Regel nicht direkt aus der Varianz von  $X$  berechnen. Eine bisweilen nützliche Approximation ist

$$\text{Var}(Y) \approx \left(g'(\mu_X)\right)^2 \text{Var}(X),$$

(approximative Formel der “Fehlerfortpflanzung”), wobei hier zur Abkürzung  $\mu_X = E(X)$  und  $g'(\mu_X)$  die Ableitung der Funktion  $g$  an der Stelle  $\mu_X$  bezeichnet. Die Näherungsformel ergibt sich aus der lokalen Linearisierung der Funktion  $g$  im Punkt  $x_0 = \mu_X$ ,

$$g(x) \approx g(\mu_X) + g'(\mu_X)(x - \mu_X) \quad \text{für alle } x,$$

so dass also näherungsweise  $Y$  als lineare Transformation von  $X$  gesehen werden kann,

$$Y = g(X) \approx g(\mu_X) + g'(\mu_X)(X - \mu_X) = aX + c$$

mit  $a = g'(\mu_X)$  und  $c = g(\mu_X) - g'(\mu_X)\mu_X$ . Die Näherungsformel für  $\text{Var}(Y)$  ergibt sich nun aus Theorem 4.15.

Wir berechnen jetzt die Varianzen für wichtige Klassen von Zufallsvariablen, die wir früher (in Kapiteln 1 und 3) eingeführt haben.

**Beispiel 4.17 (Diskrete Zufallsvariablen aus Def. 1.15 / Bsp. 4.3)**

- (a) Eine auf einer endlichen Menge  $M = \{x_1, \dots, x_m\} \subset \mathbb{R}$  gleichverteilte Zufallsvariable  $X$ :  
Wir haben gemäß Beispiel 4.3 (a):

$$E(X) = \bar{x} = \frac{1}{m} \sum_{i=1}^m x_i .$$

Mit Theorem 4.10 (a) folgt

$$\text{Var}(X) = E((X - \bar{x})^2) = \sum_{i=1}^m (x_i - \bar{x})^2 \underbrace{P(X = x_i)}_{=1/m} = \frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})^2 ,$$

und für die Standardabweichung von  $X$

$$\sqrt{\text{Var}(X)} = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})^2} .$$

Die Formel von Theorem 4.13 liefert die optisch andere Darstellung der Varianz

$$\text{Var}(X) = \frac{1}{m} \sum_{i=1}^m x_i^2 - \bar{x}^2 ,$$

denn mit Theorem 4.10(a) haben wir

$$E(X^2) = \sum_{i=1}^m x_i^2 \underbrace{P(X = x_i)}_{=1/m} = \frac{1}{m} \sum_{i=1}^m x_i^2 .$$

- (b) Eine binomial- $(n, p)$ -verteilte Zufallsvariable  $X$ :

Nach Theorem 4.10(a) ist

$$E(X^2) = \sum_{i=0}^n i^2 \binom{n}{i} p^i (1-p)^{n-i} .$$

Etwas längliche Umformungen (die wir hier nicht darstellen wollen) liefern schließlich

$$E(X^2) = n(n-1)p^2 + np .$$

Aus Beispiel 4.3(b) haben wir  $E(X) = np$ . Zusammen mit Theorem 4.13 erhalten wir

$$\text{Var}(X) = n(n-1)p^2 + np - (np)^2 = -np^2 + np = np(1-p) ,$$

und für die Standardabweichung

$$\sqrt{\text{Var}(X)} = \sqrt{np(1-p)} .$$

- (c) Eine hypergeometrisch- $(N, s, n)$ -verteilte Zufallsvariable  $X$ :  
Längliche Berechnungen ergeben

$$\text{Var}(X) = \frac{N-n}{N-1} n \frac{s}{N} \left(1 - \frac{s}{N}\right).$$

- (d) Eine Poisson- $(\lambda)$ -verteilte Zufallsvariable  $X$ :

$$\begin{aligned} \mathbb{E}(X^2) &= \sum_{i=0}^{\infty} i^2 e^{-\lambda} \frac{\lambda^i}{i!} = \lambda \sum_{i=1}^{\infty} i e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda \sum_{j=0}^{\infty} (j+1) e^{-\lambda} \frac{\lambda^j}{j!} = \lambda \left( \sum_{j=0}^{\infty} j e^{-\lambda} \frac{\lambda^j}{j!} + \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} \right) \\ &= \lambda (\mathbb{E}(X) + 1) = \lambda (\lambda + 1), \end{aligned}$$

wobei  $\mathbb{E}(X) = \lambda$  benutzt wurde (siehe Beispiel 4.3(d)). Mit Theorem 4.13 erhalten wir

$$\text{Var}(X) = \lambda(\lambda + 1) - \lambda^2 = \lambda \quad \text{bzw.} \quad \sqrt{\text{Var}(X)} = \sqrt{\lambda}.$$

- (e) Eine geometrisch- $(p)$ -verteilte Zufallsvariable  $X$ :

$$\mathbb{E}(X^2) = \sum_{i=1}^{\infty} i^2 p (1-p)^{i-1},$$

und die unendliche Summe lässt sich mit Hilfe von Resultaten über Potenzreihen explizit berechnen mit dem Ergebnis:

$$\mathbb{E}(X^2) = \frac{2(1-p)}{p^2} + \frac{1}{p}.$$

Mit Theorem 4.13 und  $\mathbb{E}(X) = 1/p$  (Beispiel 4.3(e)) ergibt dies

$$\text{Var}(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}. \quad \blacksquare$$

Sehen wir uns jetzt Varianzen von absolut-stetig-verteilten Zufallsvariablen  $X$  an, insbesondere die wichtigen Fälle aus Definition 3.6 bzw. Beispiel 4.7. Wir verwenden wieder die Formel aus Theorem 4.13,

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2,$$

sowie die Transformationsformel aus Theorem 4.10(b) für das zweite Moment,

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx,$$

wobei  $f$  die Dichtefunktion der Zufallsvariablen  $X$  bezeichnet.

#### Beispiel 4.18 (Absolut-stetig-verteilte Zufallsvariablen aus Def. 3.6 / Bsp. 4.7)

- (a) Eine normal- $(\mu, \sigma)$ -verteilte Zufallsvariable  $X$ :

Wir hatten bereits  $\mathbb{E}(X) = \mu$  gezeigt. Eine Vereinfachung der weiteren Rechnungen liefert Bemerkung 3.7: Demnach ist die Zufallsvariable

$$X_0 = \frac{X - \mu}{\sigma}$$

eine normal-(0, 1)-verteilte (d.h. eine standard-normal-verteilte) Zufallsvariable, und wir haben

$$X = \sigma X_0 + \mu .$$

Nach Theorem 4.15 ist

$$\text{Var}(X) = \sigma^2 \text{Var}(X_0) = \sigma^2 \text{E}(X_0^2) ,$$

wobei die letzte Gleichheit wegen  $\text{E}(X_0) = 0$  gilt.

Wir haben also nur noch das zweite Moment der standard-normal-verteilten Zufallsvariablen  $X_0$  zu berechnen. Die Dichtefunktion von  $X_0$  ist

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) ,$$

also

$$\text{E}(X_0^2) = \int_{-\infty}^{\infty} x^2 \phi(x) dx .$$

Das Integral lässt sich mit partieller Integration gemäß  $x^2 \phi(x) = x \cdot (x \phi(x))$  berechnen zu

$$\left[-x \phi(x)\right]_{x=-\infty}^{x=\infty} + \int_{-\infty}^{\infty} \phi(x) dx = 0 - 0 + 1 = 1 .$$

Also haben wir  $\text{E}(X_0^2) = 1$ , und daher

$$\text{Var}(X) = \sigma^2 \quad \text{bzw.} \quad \sqrt{\text{Var}(X)} = \sigma .$$

(b) Eine auf  $[\alpha, \beta]$  gleichverteilte Zufallsvariable  $X$ :

Wir hatten bereits in Beispiel 4.11 berechnet:

$$\text{E}(X^2) = \frac{1}{3} (\beta^2 + \alpha\beta + \alpha^2) .$$

Zusammen mit  $\text{E}(X) = (\alpha + \beta)/2$  erhalten wir

$$\text{Var}(X) = \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \left(\frac{\alpha + \beta}{2}\right)^2 = \frac{4\beta^2 + 4\alpha\beta + 4\alpha^2 - 3\alpha^2 - 3\beta^2 - 6\alpha\beta}{12} = \frac{(\beta - \alpha)^2}{12} .$$

Wir haben also erhalten:

$$\text{Var}(X) = \frac{(\beta - \alpha)^2}{12} \quad \text{bzw.} \quad \sqrt{\text{Var}(X)} = \frac{\beta - \alpha}{\sqrt{12}} .$$

(c) Eine exponential-( $\lambda$ )-verteilte Zufallsvariable  $X$ :

$$\text{E}(X^2) = \int_{-\infty}^{\infty} x^2 \lambda \exp(-\lambda x) \mathbf{1}_{]0, \infty[}(x) dx = \int_0^{\infty} x^2 \lambda \exp(-\lambda x) dx .$$

Letzteres Integral lässt sich mit partieller Integration berechnen zu  $2/\lambda^2$ . Wir haben also

$$\text{E}(X^2) = \frac{2}{\lambda^2} ,$$

und mit  $\text{E}(X) = 1/\lambda$  folgt

$$\text{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \quad \text{bzw.} \quad \sqrt{\text{Var}(X)} = \frac{1}{\lambda} .$$

■

Die nachfolgende Tabelle fasst noch einmal Erwartungswerte und Varianzen für die wichtigsten Klassen reeller Zufallsvariablen zusammen.

$X$	$E(X)$	$\text{Var}(X)$
0-1-wertig mit $P(X = 1) = p$	$p$	$p(1 - p)$
gleichverteilt auf $\{x_1, \dots, x_m\}$	$\bar{x}$	$\frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})^2$
binomial- $(n, p)$ -verteilt	$np$	$np(1 - p)$
hypergeometrisch- $(N, s, n)$ -verteilt	$\frac{ns}{N}$	$\frac{N - n}{N - 1} n \frac{s}{N} \left(1 - \frac{s}{N}\right)$
Poisson- $(\lambda)$ -verteilt	$\lambda$	$\lambda$
geometrisch- $(p)$ -verteilt	$\frac{1}{p}$	$\frac{1 - p}{p^2}$
normal- $(\mu, \sigma)$ -verteilt	$\mu$	$\sigma^2$
gleichverteilt auf $[\alpha, \beta]$	$\frac{1}{2}(\beta + \alpha)$	$\frac{(\beta - \alpha)^2}{12}$
exponential- $(\lambda)$ -verteilt	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Tabelle 4.1

### 4.3 Varianzen von Linearkombinationen

Seien  $X$  und  $Y$  zwei reelle Zufallsvariablen auf  $\Omega$  und  $a, b$  und  $c$  reelle Zahlen. Wir betrachten die Linearkombination

$$Z = aX + bY + c, \quad \text{d.h.} \quad Z(\omega) = aX(\omega) + bY(\omega) + c \quad \text{für alle } \omega \in \Omega,$$

die wiederum eine reelle Zufallsvariable auf  $\Omega$  ist, und deren Varianz. Zunächst haben wir für deren Erwartungswert (nach Theorem 4.8)

$$E(Z) = aE(X) + bE(Y) + c.$$

Folglich ist

$$Z - E(Z) = a(X - E(X)) + b(Y - E(Y)),$$

und das Quadrat hiervon:

$$(Z - E(Z))^2 = a^2(X - E(X))^2 + b^2(Y - E(Y))^2 + 2ab[X - E(X)][Y - E(Y)];$$

die Bildung des Erwartungswertes liefert (unter Verwendung von Definition 4.12 und Theorem 4.8):

$$\text{Var}(Z) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y),$$

wobei wir bezeichnet haben:

$$\text{Cov}(X, Y) = E\left([X - E(X)][Y - E(Y)]\right),$$

die sog. *Kovarianz* der beiden Zufallsvariablen  $X$  und  $Y$ . Nur wenn diese gleich Null ist, lässt sich also die Varianz der Linearkombination  $Z$  mit Hilfe der einzelnen Varianzen (der von  $X$  und von  $Y$ ) berechnen. Ohne weitere Erläuterung sei hier mitgeteilt, dass im Falle der stochastischen Unabhängigkeit der beiden Zufallsvariablen  $X$  und  $Y$  ihre Kovarianz gleich Null ist. Wir erhalten damit:

**Theorem 4.19 (Varianz einer LK zweier stoch. unabh. reeller ZV'en)**

Für stochastisch unabhängige reelle Zufallsvariablen  $X$  und  $Y$  und reelle Zahlen  $a, b, c$  gilt:

$$\text{Var}(aX + bY + c) = a^2 \text{Var}(X) + b^2 \text{Var}(Y).$$

Eine Erweiterung auf Linearkombinationen mehrerer Zufallsvariablen formuliert das folgende Theorem.

**Theorem 4.20**

Für  $n \geq 2$  stochastisch unabhängige reelle Zufallsvariablen  $X_1, X_2, \dots, X_n$  und reelle Zahlen  $a_1, a_2, \dots, a_n$  und  $c$  gilt:

$$\text{Var}\left(\sum_{i=1}^n a_i X_i + c\right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i).$$

Manchmal sind nicht nur der Erwartungswert oder die Varianz einer Linearkombination von Zufallsvariablen von Interesse, sondern darüberhinaus die Verteilung der Linearkombination. Im Fall stochastisch unabhängiger und normal-verteilter Zufallsvariablen  $X_1, \dots, X_n$  lässt sich Folgendes aussagen.

**Theorem 4.21 (Linearkomb. stoch. unabh. normal-verteilter Zufallsvariablen)**

Seien  $X_1, X_2, \dots, X_n$  stoch. unabhängige reelle Zufallsvariablen, und  $X_i$  sei normal- $(\mu_i, \sigma_i)$ -verteilt für jedes  $i = 1, 2, \dots, n$ . Seien  $a_1, a_2, \dots, a_n$  und  $c$  reelle Zahlen, wobei  $a_i \neq 0$  für mindestens ein  $i$  sei. Betrachte die Linearkombination

$$Z = \sum_{i=1}^n a_i X_i + c.$$

Dann ist  $Z$  eine normal- $(\mu_Z, \sigma_Z)$ -verteilte Zufallsvariable, wobei

$$\mu_Z = \sum_{i=1}^n a_i \mu_i + c \quad \text{und} \quad \sigma_Z^2 = \sum_{i=1}^n a_i^2 \sigma_i^2.$$

## Kapitel 5

# Statistische Modelle und Fragestellungen

### 5.1 Statistische Modellierung

Bei der Modellierung eines Zufallsexperiments besteht oft (wenn nicht sogar meistens) Unsicherheit darüber, von welcher Verteilung der Zufallsvariablen ausgegangen werden sollte. Das in Abschnitt 1.1 dargestellte Konzept setzt voraus, dass die Verteilung der Zufallsvariablen vollständig und eindeutig festgelegt wird. Das ist aber zum Beispiel schon für das simple Zufallsexperiment des Münzwurfes (einmal, oder auch  $n$ -mal unabhängig) dann nicht mehr gegeben, wenn Zweifel an der idealen Beschaffenheit der Münze angemeldet werden, d. h. wenn die Möglichkeit in Betracht gezogen wird, dass die Wahrscheinlichkeiten für “1” und “0” nicht jeweils  $1/2$  betragen, sondern  $p$  und  $1 - p$ , wobei Unsicherheit über den wahren Wert  $p \in ]0, 1[$  besteht. Das mag für den Münzwurf etwas weit her geholt erscheinen; ersetzen wir aber den Münzwurf durch ein anderes 0-1-Experiment (z. B. Erfolg/Misserfolg einer Therapiemaßnahme in einer klinischen Studie), so wird die Unsicherheit über den Wert  $p$  (Erfolgswahrscheinlichkeit) eher die Regel sein. Oder denken wir an ein Zufallsexperiment, das “kontinuierliche” (reellwertige) Ergebnisse hat, wie z. B. Produkt-Charakteristika in der industriellen Fertigung. Hier verwendet man zur Modellierung oft und gerne eine (oder mehrere unabhängige) normalverteilte Zufallsvariable(n). Abgesehen davon, dass diese Annahme grundsätzlich angezweifelt werden kann (warum normal-verteilt?), ergibt sich die Frage, welche Parameterwerte  $\mu$  und  $\sigma$  verwendet werden sollten. Die statistische Modellierung berücksichtigt die Unsicherheit über die Verteilungsannahme zumindest innerhalb eines gewissen Rahmens: Es werden einige Parameter in das Verteilungsmodell aufgenommen, deren Werte offen gelassen werden, also unbekannt sind. Statistische Modellierungen können in solchen Situationen so aussehen:

#### Beispiel 5.1 (0-1-Experimente)

- (a) Ein 0-1-Experiment werde  $n$ -mal unabhängig durchgeführt; dabei ist die Wahrscheinlichkeit für “1” im Einzelexperiment unbekannt. Die Modellierung erfolgt durch  $n$  0-1-wertige Zufallsvariablen  $X_1, \dots, X_n$ , die stochastisch unabhängig sind und für die gilt:

$$P(X_i = 1) = p \quad \text{für alle } i = 1, \dots, n.$$

Der Wert  $p \in ]0, 1[$  ist unbekannt;  $p$  ist also der Parameter im Modell.

- (b) Als Ergebnis des Zufallsexperiments werde jetzt nur die Anzahl der in den  $n$  Einzelexperimenten erzielten “1”en (z. B. Anzahl der Erfolge) notiert. Diese wird als eine Zufallsvariable

$X$  mit Werten in der Menge  $\{0, 1, \dots, n\}$  aufgefasst, und gemäß Beispiel 2.11 wird diese als binomial-( $n, p$ )-verteilt angenommen. Der Wert  $p \in ]0, 1[$  ist aber unbekannt; wiederum ist also  $p$  der Parameter im Modell.

Beachte: Die Anzahl  $n$  der Durchführungen des zu Grunde liegenden 0-1-Experiments ist hier (in (a) und (b)) natürlich als bekannt vorausgesetzt (z. B.  $n = 20$ ).

### Beispiel 5.2 (Ein Normalverteilungsmodell)

Es werden die Werte einer kontinuierlichen Produkt-Charakteristik von  $n$  gefertigten (gleichartigen) Produkten gemessen. Die zufällige (oder zufällig erscheinende) Schwankung der Werte legt es nahe, diese als Werte von Zufallsvariablen  $X_1, \dots, X_n$  aufzufassen. Mit Blick auf die üblichen und (zumindest teilweise) bewährten Modellansätze treffen wir folgende Annahmen:

Die Zufallsvariablen sind stochastisch unabhängig und identisch-verteilt (u.i.v.), und spezieller: sie sind (identisch) normal-( $\mu, \sigma$ )-verteilt. Die Werte für  $\mu \in \mathbb{R}$  und  $\sigma > 0$  sind aber unbekannt. Wir haben also zwei Parameter in diesem Modell,  $\mu$  und  $\sigma$ , oder anders gesagt: wir haben den zwei-dimensionalen Parameter  $(\mu, \sigma)$ . ■

Grundsätzlich können wir festhalten: Die statistische Modellierung eines Zufallsexperiments gestaltet die eingangs in Abschnitt 1.1 beschriebene Modellierung flexibler, indem die Verteilung der Zufallsvariablen nicht vollständig festgelegt wird, sondern eine “parametrische” Familie möglicher Verteilungen angesetzt wird. Es wird aber davon ausgegangen, dass die wahre Verteilung der Zufallsvariablen zu dieser Familie gehört. Das mag zwar immer noch restriktiv (bisweilen vielleicht auch willkürlich) erscheinen, aber immerhin: Es handelt sich um eine schwächere Annahme als es eine eindeutige Festlegung auf eine Verteilung (d. h. auf einen speziellen Wert des Parameters) bedeuten würde. Wir merken noch an, dass unsere Sprechweise *ein Parameter* im Allgemeinen auch die Möglichkeit beinhalten soll, dass der Parameter *mehrdimensional* ist, also praktisch mehrere (eindimensionale) Parameter vorhanden sind (in Beispiel 5.2 haben wir einen zwei-dimensionalen Parameter  $(\mu, \sigma)$  bzw. die beiden Parameter  $\mu$  und  $\sigma$ ).

Wir wollen nun diese grundsätzliche Vorgehensweise etwas formaler beschreiben mit Blick auf die zu verwendenden Begriffe *W-Raum*, *Zufallsvariable* und ihre *Verteilung*. Zunächst gehen wir – wie bisher – von einem W-Raum  $(\Omega, P)$  aus, der meist nicht weiter spezifiziert wird. Ein Ergebnis des zu modellierenden Zufallsexperiments fassen wir als Wert einer Zufallsvariablen  $X : \Omega \rightarrow M$  auf, wobei  $M$  die Menge aller möglichen Ergebnisse des Zufallsexperiments bezeichnet. Die Verteilung der Zufallsvariablen  $X$  ist nun allgemein durch die Wahrscheinlichkeiten

$$P(X \in C) \quad \text{für alle } C \subset M \quad (5.1)$$

gegeben.

Diese sollen jetzt – im statistischen Kontext – noch einen Parameter  $\vartheta \in \Theta$  beinhalten, wobei dieser auch mehrdimensional sein kann; der wahre Wert des Parameters  $\vartheta$  ist (innerhalb eines festgelegten Parameterbereiches  $\Theta$ ) unbekannt. Allerdings geraten wir mit dieser Sichtweise zunächst in einen formalen Konflikt mit dem Begriff der *Verteilung der Zufallsvariablen*  $X$ , die durch (5.1) gegeben ist. Für einen W-Raum  $(\Omega, P)$  und für eine Zufallsvariable  $X$  auf  $\Omega$  ist durch (5.1) ja die Verteilung der Zufallsvariablen  $X$  gegeben, und nicht eine ganze Familie möglicher Verteilungen, die man durch Variation des Wertes von  $\vartheta \in \Theta$  erhalten würde. Der Ausweg kann nur sein:

Wir gehen auch von einer *Familie von W-Verteilungen* auf  $\Omega$  aus; also statt einer W-Verteilung  $P$  auf  $\Omega$  stellen wir uns vor, dass eine Familie von W-Verteilungen auf  $\Omega$  vorhanden ist,

$$P_{\vartheta} \quad \text{für alle } \vartheta \in \Theta.$$

Eine dieser Verteilungen ist die wahre, aber wir wissen nicht welche (wir kennen den wahren Wert des Parameters  $\vartheta$  nicht). Mit dieser Sichtweise schreibt sich (5.1) besser als

$$P_{\vartheta}(X \in C) \quad \text{für alle } C \subset M \text{ und alle } \vartheta \in \Theta, \quad (5.2)$$



womit der formale Konflikt beseitigt ist: Es werden ja in (5.2) die Verteilungen der Zufallsvariablen  $X : \Omega \longrightarrow M$  unter den verschiedenen W-Verteilungen  $P_\vartheta$  auf  $\Omega$  angesprochen und nicht die Verteilung von  $X$  unter einer W-Verteilung  $P$  auf  $\Omega$ . In diesem Zusammenhang ist daran zu erinnern, dass die Verteilung einer Zufallsvariablen gemäß Definition 1.11 aus der Zufallsvariablen und der zu Grunde gelegten W-Verteilung auf  $\Omega$  gebildet wird und daher bei Variation letzterer eine andere wird.

Oft haben wir es mit Zufallsexperimenten zu tun, die aus  $n \geq 2$  voneinander unabhängigen Einzelexperimenten bestehen. Die Modellierung erfolgt dann durch  $n$  Zufallsvariablen  $X_1, \dots, X_n$ , wobei

$$X_i : \Omega \longrightarrow M_i, \quad i = 1, \dots, n,$$

das Ergebnis des  $i$ -ten Einzelperiments repräsentiert. Die Unabhängigkeit der Einzelexperimente wird durch die Annahme erfasst, dass die Zufallsvariablen  $X_1, \dots, X_n$  stochastisch unabhängig sind. Da wir im statistischen Kontext einen Parameter  $\vartheta \in \Theta$  haben, wir also von einer Familie  $P_\vartheta$  ( $\vartheta \in \Theta$ ) von W-Verteilungen auf  $\Omega$  ausgehen müssen, bedeutet die stochastische Unabhängigkeit der Zufallsvariablen  $X_1, \dots, X_n$  präziser formuliert:

*Für jedes  $\vartheta \in \Theta$  sind die Zufallsvariablen  $X_1, \dots, X_n$  unter  $P_\vartheta$  stochastisch unabhängig, ihre gemeinsame Verteilung (unter  $P_\vartheta$ ) ist also das Produkt ihrer einzelnen Verteilungen (unter  $P_\vartheta$ ):*

$$P_\vartheta(X_1 \in C_1, X_2 \in C_2, \dots, X_n \in C_n) = P_\vartheta(X_1 \in C_1) \cdot P_\vartheta(X_2 \in C_2) \cdots P_\vartheta(X_n \in C_n)$$

*für alle  $C_1 \subset M_1, C_2 \subset M_2, \dots, C_n \subset M_n$ , und für alle  $\vartheta \in \Theta$  (vgl. Abschnitt 2.3).*

Dabei sind die einzelnen Verteilungen

$$P_\vartheta(X_i \in C_i) \quad \text{für alle } C_i \subset M_i \text{ und alle } \vartheta \in \Theta,$$

für jedes  $i = 1, \dots, n$  entsprechend der konkreten Situation geeignet festzulegen.

Oft gehen wir von obendrein *identisch verteilten* Zufallsvariablen aus (dann also von u.i.v. Zufallsvariablen  $X_1, \dots, X_n$ ), d. h. die einzelnen Verteilungen sind dann für jedes  $\vartheta \in \Theta$  identisch (sie hängen also nicht von  $i$ , sondern nur von  $\vartheta$  ab). In den obigen Beispielen 5.1(a) und 5.2 haben wir solche Situationen:

In Beispiel 5.1(a): Die 0-1-wertigen Zufallsvariablen  $X_1, \dots, X_n$  sind u.i.v. (unter  $P_p$ , für jedes  $p \in ]0, 1[$ ) mit

$$P_p(X_i = 1) = p \quad (\text{und } P_p(X_i = 0) = 1 - p).$$

In Beispiel 5.2: Die reellen Zufallsvariablen  $X_1, \dots, X_n$  sind u.i.v. normal- $(\mu, \sigma)$ -verteilt (unter  $P_{\mu, \sigma}$ , für jedes  $\mu \in \mathbb{R}$  und jedes  $\sigma > 0$ ).

Im Folgenden werden wir bei den Formulierungen der diversen statistischen Modelle die Zusätze “unter  $P_\vartheta$ ”, “für jedes  $\vartheta \in \Theta$ ” meistens nicht mehr explizit aufnehmen; wohl aber wollen wir stets die zu Grunde gelegten W-Verteilungen auf  $\Omega$  als  $P_\vartheta$  schreiben, damit die Präsenz des Parameters  $\vartheta \in \Theta$  in den Verteilungen der Zufallsvariablen auch formal zum Ausdruck gebracht wird.

## 5.2 Spezielle Modelle

Bei der statistischen Modellierung haben wir es in der Regel entweder mit *diskreten* Zufallsvariablen oder mit *absolut-stetig*-verteilten Zufallsvariablen zu tun. Die Beschreibung der möglichen Verteilungen erfolgt prinzipiell so:

- **Im diskreten Fall**

Wenn wir eine diskrete Zufallsvariable  $X : \Omega \longrightarrow M$  im Modell haben, dann ist eine Formel bereitzustellen für die Wahrscheinlichkeiten

$$P_{\vartheta}(X = x) \quad \text{für alle } x \in M \text{ und alle } \vartheta \in \Theta.$$

Wenn wir  $n \geq 2$  diskrete Zufallsvariablen  $X_1, X_2, \dots, X_n$  im Modell haben und diese als stochastisch unabhängig vorausgesetzt werden, dann ist eine Formel bereitzustellen für die Produkt-Wahrscheinlichkeiten

$$P_{\vartheta}(X_1 = x_1) \cdot P_{\vartheta}(X_2 = x_2) \cdot \dots \cdot P_{\vartheta}(X_n = x_n) \\ \text{für alle } x_1 \in M_1, x_2 \in M_2, \dots, x_n \in M_n \text{ und alle } \vartheta \in \Theta.$$

- **Im absolut-stetigen Fall**

Wenn wir eine reelle (absolut-stetig-verteilte) Zufallsvariable  $X : \Omega \longrightarrow \mathbb{R}$  im Modell haben, dann ist eine Formel für die *Dichtefunktionen* bereitzustellen,

$$f(x; \vartheta) \quad \text{für alle } x \in \mathbb{R} \text{ und alle } \vartheta \in \Theta.$$

Beachte: Die Schreibweise  $f(x; \vartheta)$  steht hier für die Dichtefunktion der Zufallsvariablen  $X$  unter der  $W$ -Verteilung  $P_{\vartheta}$ . Die Verteilung und daher auch die Dichtefunktion der Zufallsvariablen  $X$  hängen ja vom Parameter  $\vartheta$  ab, und eben dies wird durch die Schreibweise zum Ausdruck gebracht.

Wenn wir  $n \geq 2$  reelle (absolut-stetig-verteilte) Zufallsvariablen  $X_1, X_2, \dots, X_n$  im Modell haben und diese als stochastisch unabhängig vorausgesetzt werden, dann ist eine Formel für Produkt-Dichtefunktionen bereitzustellen,

$$f_{X_1}(x_1; \vartheta) \cdot f_{X_2}(x_2; \vartheta) \cdot \dots \cdot f_{X_n}(x_n; \vartheta) \\ \text{für alle } x_1, x_2, \dots, x_n \in \mathbb{R} \text{ und alle } \vartheta \in \Theta,$$

wobei  $f_{X_i}(x_i; \vartheta)$  die Dichtefunktion der Zufallsvariablen  $X_i$  ist, für jedes  $i = 1, 2, \dots, n$ .

Es sei angemerkt, dass die Verwendung der Produkt-Dichtefunktion hier nicht weiter begründet werden kann; sie erscheint jedenfalls plausibel – als stetiges Analogon zum diskreten Fall.

### Beispiel 5.3 (Binomialmodell)

Betrachten wir das in Beispiel 5.1(b) angesprochene Modell:

Eine binomial- $(n, p)$ -verteilte Zufallsvariable, wobei  $p \in ]0, 1[$  unbekannt ist. Hier heißt  $\vartheta$  also  $p$  und der Parameterbereich  $\Theta$  ist das Intervall  $]0, 1[$ . Da  $X$  diskret ist (die möglichen Werte von  $X$  sind  $0, 1, \dots, n$ ), sind die möglichen Verteilungen von  $X$  gegeben durch

$$P_p(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad \text{für alle } x = 0, 1, \dots, n \text{ und für alle } p \in ]0, 1[. \quad \blacksquare$$

### Beispiel 5.4 (Hypergeometrisches Modell)

Gegeben sei eine Grundgesamtheit von  $N$  Objekten, wobei jedes Objekt einen binären Merkmalswert “0” oder “1” trägt (z.B. “gut” oder “schlecht”). Insgesamt sind  $s$  Objekte mit Merkmalswert “1” vorhanden, diese Anzahl ist aber unbekannt. Eine Zufallsstichprobe von  $n$  Objekten wird aus der Grundgesamtheit gezogen (“ohne Zurücklegen”), und die Anzahl  $x$  der Objekte mit Wert “1” in der Stichprobe wird festgestellt (diese Anzahl bildet das Ergebnis des Zufallsexperiments). Beachte: die Größe  $N$  der Grundgesamtheit wie auch die Stichprobengöße  $n$  sind bekannt, während die Anzahl  $s$  der “1”-er Objekte in der Grundgesamtheit unbekannt ist. In Beispiel 2.18 haben wir

gesehen, dass für die beobachtete Anzahl  $x$  eine hypergeometrische Verteilung zu Grunde zu legen ist.

Das statistische Modell ist daher: Beobachtet wird ein Wert einer hypergeometrisch- $(N, s, n)$ -verteilten Zufallsvariablen  $X : \Omega \longrightarrow \{0, 1, \dots, n\}$ , wobei  $s \in \{0, 1, \dots, N\}$  den unbekannten Parameter darstellt; formelmäßig ausgedrückt:

$$P_s(X = x) = \frac{\binom{s}{x} \binom{N-s}{n-x}}{\binom{N}{n}} \quad \text{für alle } x = 0, 1, \dots, n \text{ und alle } s = 0, 1, \dots, N.$$

### Beispiel 5.5 (Poisson-Modell)

- (a) Eine zufällige Anzahl  $x$  wird beobachtet (z. B. Anzahl der Kunden einer Service-Station in einem definierten Zeitraum). Wir nehmen an, dass diese Anzahl der Wert einer Poisson-verteilten Zufallsvariablen ist, wobei aber der wahre Wert für  $\lambda > 0$  der Poisson-Verteilung unbekannt ist.

Das statistische Modell lautet: Beobachtet wird der Wert einer Poisson- $(\lambda)$ -verteilten Zufallsvariablen  $X : \Omega \longrightarrow \mathbb{N}_0$ , und  $\lambda > 0$  ist der Parameter. Formelmäßig ausgedrückt:

$$P_\lambda(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad \text{für alle } x \in \mathbb{N}_0 \text{ und alle } \lambda > 0.$$

- (b) Jetzt werden wiederholt (unabhängig voneinander) zufällige Anzahlen  $x_1, x_2, \dots, x_n$  beobachtet (wobei  $n \geq 2$  gegeben ist), z. B. die Anzahlen der Kunden an verschiedenen Tagen (aber jeweils über gleich lange Zeiträume hinweg). Wir fassen diese Anzahlen als Werte von u.i.v. Zufallsvariablen  $X_1, X_2, \dots, X_n$  auf, wobei die Zufallsvariablen identisch Poisson- $(\lambda)$ -verteilt mit unbekanntem  $\lambda > 0$  sind.

Das statistische Modell besteht also aus  $n$  stochastisch unabhängigen und identisch Poisson- $(\lambda)$ -verteilten Zufallsvariablen  $X_i : \Omega \longrightarrow \mathbb{N}_0$ ,  $i = 1, 2, \dots, n$ , d. h. formelmäßig:

$$\begin{aligned} & P_\lambda(X_1 = x_1) \cdot P_\lambda(X_2 = x_2) \cdots P_\lambda(X_n = x_n) \\ &= e^{-\lambda} \frac{\lambda^{x_1}}{x_1!} \cdot e^{-\lambda} \frac{\lambda^{x_2}}{x_2!} \cdots e^{-\lambda} \frac{\lambda^{x_n}}{x_n!} = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{x_1! x_2! \cdots x_n!} \\ & \text{für alle } x_1, x_2, \dots, x_n \in \mathbb{N}_0 \text{ und alle } \lambda > 0. \end{aligned}$$

### Beispiel 5.6 (Geometrisches Modell)

- (a) Eine zufällige diskrete Zeit  $x \in \mathbb{N}$  wird beobachtet (z. B. eine Wartezeit), die als geometrisch- $(p)$ -verteilt angenommen wird mit einem unbekannten Wert  $p \in ]0, 1[$ .

Wir haben also das Modell einer geometrisch- $(p)$ -verteilten Zufallsvariablen  $X : \Omega \longrightarrow \mathbb{N}$ , wobei  $p$  der Parameter ist; formelmäßig:

$$P_p(X = x) = p(1-p)^{x-1} \quad \text{für alle } x \in \mathbb{N} \text{ und alle } p \in ]0, 1[.$$

- (b) Jetzt wird das Zufallsexperiment in (a)  $n \geq 2$  mal unabhängig durchgeführt, also z. B. Wartezeiten in  $n$  gleichartigen Situationen beobachtet (wobei die Unabhängigkeit unterstellt wird). Als Modell nehmen wir  $n$  stochastisch unabhängige und identisch geometrisch- $(p)$ -verteilte Zufallsvariablen  $X_1, X_2, \dots, X_n$ , und  $p \in ]0, 1[$  ist der Parameter. Formelmäßig:

$$\begin{aligned} & P_p(X_1 = x_1) \cdot P_p(X_2 = x_2) \cdots P_p(X_n = x_n) \\ &= p(1-p)^{x_1-1} \cdot p(1-p)^{x_2-1} \cdots p(1-p)^{x_n-1} = p^n (1-p)^{\sum_{i=1}^n x_i - n} \\ & \text{für alle } x_1, x_2, \dots, x_n \in \mathbb{N} \text{ und alle } p \in ]0, 1[. \end{aligned}$$

**Beispiel 5.7 (Exponentialverteilungsmodell)**

- (a) Beobachtet wird eine zufällige “kontinuierliche” Zeitdauer (z. B. eine Lebensdauer, eine Wartezeit). Wir nehmen an, dass diese exponential- $(\lambda)$ -verteilt ist mit einem unbekannten  $\lambda > 0$ . Das Modell besteht also aus einer exponential- $(\lambda)$ -verteilten reellen Zufallsvariablen  $X : \Omega \rightarrow \mathbb{R}$ , wobei  $\lambda$  der Parameter ist. Eine formelmäßige Beschreibung des Modells erfolgt durch die Dichtefunktion von  $X$  (unter  $P_\lambda$  für jedes  $\lambda > 0$ ):

$$f(x; \lambda) = \begin{cases} \lambda \exp(-\lambda x) & , \text{ falls } x > 0 \\ 0 & , \text{ sonst} \end{cases} \quad \text{für alle } x \in \mathbb{R} \text{ und alle } \lambda > 0.$$

- (b) Beobachtet werden jetzt  $n \geq 2$  zufällige “kontinuierliche” Zeitdauern (z. B. Lebensdauern gleichartiger Objekte). Wir nehmen an, dass diese unabhängig und identisch exponential- $(\lambda)$ -verteilt sind mit einem unbekannten  $\lambda > 0$ . Das Modell besteht also aus  $n$  stochastisch unabhängigen und identisch exponential- $(\lambda)$ -verteilten reellen Zufallsvariablen  $X_1, X_2, \dots, X_n$ , und  $\lambda > 0$  ist der Parameter. Eine formelmäßige Beschreibung erfolgt durch die Produkt-Dichtefunktion

$$f_{X_1}(x_1; \lambda) \cdot f_{X_2}(x_2; \lambda) \cdot \dots \cdot f_{X_n}(x_n; \lambda) ,$$

wobei  $f_{X_i}(x_i; \lambda)$  die Dichtefunktion der (einzelnen) Zufallsvariablen  $X_i$  ist ( $i = 1, 2, \dots, n$ ), d. h. (da jede Zufallsvariable  $X_i$  exponential- $(\lambda)$ -verteilt ist)

$$f_{X_i}(x_i; \lambda) = \begin{cases} \lambda \exp(-\lambda x_i) & , \text{ falls } x_i > 0 \\ 0 & , \text{ sonst} \end{cases} .$$

Wir erhalten für die Produkt-Dichtefunktion:

$$\begin{aligned} & \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right) & , \text{ falls } x_1, x_2, \dots, x_n > 0 \\ & 0 & , \text{ sonst} \end{aligned}$$

für alle  $x_1, x_2, \dots, x_n \in \mathbb{R}$  und alle  $\lambda > 0$ .

**Beispiel 5.8 (Normalverteilungsmodell, vgl. Beispiel 5.2)**

- (a) Modell mit einer normal- $(\mu, \sigma)$ -verteilten Zufallsvariablen  $X$  (beschreibt z. B. den Wert einer Produkt-Charakteristik eines gefertigten Teils, das zufällig der Produktion entnommen wird).

Wir haben den zwei-dimensionalen Parameter  $\vartheta = (\mu, \sigma)$ ; die Dichtefunktionen der Zufallsvariablen  $X$  sind:

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right) \quad \text{für alle } x \in \mathbb{R} \text{ und alle } \mu \in \mathbb{R}, \sigma > 0.$$

- (b) Modell mit  $n \geq 2$  u.i.v. normal- $(\mu, \sigma)$ -verteilten Zufallsvariablen  $X_1, X_2, \dots, X_n$  (beschreiben z. B. die Werte einer Produkt-Charakteristik von  $n$  gefertigten Teilen, die zufällig der Produktion entnommen werden).

Wir haben wiederum den zwei-dimensionalen Parameter  $\vartheta = (\mu, \sigma)$ ; die Dichtefunktion der einzelnen Zufallsvariablen  $X_i$  ist (da jede Zufallsvariable  $X_i$  als normal- $(\mu, \sigma)$ -verteilt vorausgesetzt wird),

$$f_{X_i}(x_i; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{x_i-\mu}{\sigma}\right]^2\right) ,$$

für jedes  $i = 1, 2, \dots, n$ . Die Produkt-Dichte ist daher (nach offensichtlichen Umformungen):

$$f_{X_1}(x_1; \mu, \sigma) \cdot f_{X_2}(x_2; \mu, \sigma) \cdot \dots \cdot f_{X_n}(x_n; \mu, \sigma) = \frac{1}{\sigma^n(\sqrt{2\pi})^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

für alle  $x_1, x_2, \dots, x_n \in \mathbb{R}$  und alle  $\mu \in \mathbb{R}, \sigma > 0$ .

**Beispiel 5.9 (Zwei Stichproben: Normalverteilungsmodell)**

Zwei verschiedene Fertigungsmethoden sollen hinsichtlich einer quantitativen Produkt-Charakteristik verglichen werden. Aus der Fertigung unter Methode 1 wird eine Stichprobe vom Umfang  $n_1$  entnommen und aus der Fertigung unter Methode 2 eine Stichprobe vom Umfang  $n_2$ . Die beobachteten Werte der Produkt-Charakteristik seien mit  $x_1, \dots, x_{n_1}$  und  $y_1, \dots, y_{n_2}$  bezeichnet. Die Schwankungen der Werte innerhalb der beiden Stichproben werden auf Zufallseinflüsse zurückgeführt. Ein oft verwendetes Modell ist das folgende Normalverteilungsmodell:

Die Beobachtungswerte werden als Werte von reellen Zufallsvariablen

$$X_1, \dots, X_{n_1} \quad \text{und} \quad Y_1, \dots, Y_{n_2}$$

aufgefasst. Alle  $n = n_1 + n_2$  Zufallsvariablen  $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$  sind stochastisch unabhängig. Die Zufallsvariablen  $X_1, \dots, X_{n_1}$  sind identisch normal- $(\mu_1, \sigma)$ -verteilt, und die Zufallsvariablen  $Y_1, \dots, Y_{n_2}$  sind identisch normal- $(\mu_2, \sigma)$ -verteilt. Dabei sind  $\mu_1, \mu_2 \in \mathbb{R}$  und  $\sigma > 0$  unbekannt. Das Modell beinhaltet also den drei-dimensionalen Parameter  $\vartheta = (\mu_1, \mu_2, \sigma)$ .

Beachte: Die Modellannahmen beinhalten insbesondere, dass für die beiden Fertigungsmethoden möglicherweise unterschiedliche Werte für den  $\mu$ -Parameter ( $\mu_1$  und  $\mu_2$ ), aber derselbe Wert für den  $\sigma$ -Parameter der normal-verteilten Zufallsvariablen unterstellt wird.

Die Dichtefunktionen der einzelnen Zufallsvariablen  $X_i$  und  $Y_j$  sind gegeben durch

$$\begin{aligned} f_{X_i}(x_i; \mu_1, \sigma) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{x_i - \mu_1}{\sigma}\right]^2\right) \quad \text{für alle } x_i \in \mathbb{R} \text{ und alle } \mu_1 \in \mathbb{R}, \sigma > 0, \\ f_{Y_j}(y_j; \mu_2, \sigma) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{y_j - \mu_2}{\sigma}\right]^2\right) \quad \text{für alle } y_j \in \mathbb{R} \text{ und alle } \mu_2 \in \mathbb{R}, \sigma > 0. \end{aligned}$$

Für die Produkt-Dichte erhalten wir also nach offensichtlichen Umformungen):

$$\begin{aligned} &f_{X_1}(x_1; \mu_1, \sigma) \cdot \dots \cdot f_{X_{n_1}}(x_{n_1}; \mu_1, \sigma) \cdot f_{Y_1}(y_1; \mu_2, \sigma) \cdot \dots \cdot f_{Y_{n_2}}(y_{n_2}; \mu_2, \sigma) \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_1}(x_i - \mu_1)^2 + \sum_{j=1}^{n_2}(y_j - \mu_2)^2\right]\right) \\ &\quad \text{für alle } x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2} \in \mathbb{R} \text{ und alle } \mu_1, \mu_2 \in \mathbb{R}, \sigma > 0. \end{aligned}$$

**Beispiel 5.10 (Regressionsgerade: Normalverteilungsmodell)**

Eine reelle Zielgröße  $y$  hänge von einer reellen Einflussgröße  $t$  ab:  $y = y(t)$ . Z. B. ist  $t$  die Konzentration eines Stoffes in einer Lösung und  $y = y(t)$  die daraus resultierende elektrische Leitfähigkeit der Lösung. In erster Näherung wird davon ausgegangen, dass eine *lineare* Abhängigkeit besteht, d. h.

$$y(t) = at + c$$

für alle möglichen Werte  $t$  innerhalb eines gewissen Bereichs (in der Regel ein Intervall), wobei  $a$  und  $c$  reelle Konstanten sind, die aber unbekannt sind. Es liegen nun Beobachtungsdaten vor:

$$(t_1, y_1), (t_2, y_2), \dots, (t_n, y_n),$$

z. B. wurde jeweils bei einer Konzentration  $t_i$  der Signalwert (elektrische Leitfähigkeit)  $y_i$  beobachtet ( $i = 1, \dots, n$ ). Die  $y_i$ -Werte hängen von den jeweiligen  $t_i$ -Werten ab, aber außerdem noch von unkontrollierbaren Zufallseinflüssen. Daher wird die folgende *statistische* Sichtweise in das Modell eingebracht: Die beobachteten Werte  $y_1, \dots, y_n$  werden als Werte von Zufallsvariablen  $Y_1, Y_2, \dots, Y_n$  interpretiert. Diese Zufallsvariablen werden als stochastisch unabhängig und normalverteilt angenommen, und zwar  $Y_i$  als normal- $(y(t_i), \sigma)$ -verteilt ( $i = 1, \dots, n$ ), wobei gemäß der eingangs getroffenen Annahme einer linearen Funktion  $y(t)$ ,

$$y(t_i) = at_i + c \quad \text{für } i = 1, \dots, n.$$

Als unbekannte Parameter im Modell haben wir  $a, c \in \mathbb{R}$  und  $\sigma > 0$ .

Beachte: Nur die beobachteten Werte  $y_1, \dots, y_n$  werden hier als Werte von Zufallsvariablen behandelt, nicht aber die beobachteten Werte  $t_1, \dots, t_n$ . Letztere werden wie bekannte Konstanten behandelt (sie unterliegen keinen Zufallseinflüssen).

Eine explizitere Modellformulierung erfolgt durch die Produkt-Dichtefunktion der Zufallsvariablen  $Y_1, \dots, Y_n$ :

Zunächst haben wir den drei-dimensionalen Parameter  $\vartheta = (a, c, \sigma)$ , wobei  $a, c \in \mathbb{R}$  und  $\sigma > 0$ . Jede *einzelne* Zufallsvariable  $Y_i$  ist normal- $(at_i + c, \sigma)$ -verteilt (unter  $P_\vartheta$ ) besitzt also die Dichtefunktion

$$f_{Y_i}(y_i; a, c, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{y_i - (at_i + c)}{\sigma}\right]^2\right) \quad \text{für alle } y_i \in \mathbb{R},$$

für jedes  $i = 1, \dots, n$  (und jedes  $\vartheta = (a, c, \sigma)$ ).

Beachte: Die Zufallsvariablen  $Y_1, \dots, Y_n$  sind hier nicht identisch verteilt.

Wohl aber sind die Zufallsvariablen  $Y_1, \dots, Y_n$  als stochastisch unabhängig vorausgesetzt, so dass wir als Produkt-Dichtefunktion erhalten (nach offensichtlichen Umformungen):

$$\prod_{i=1}^n f_{Y_i}(y_i; a, c, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - at_i - c)^2\right].$$

### 5.3 Statistische Fragestellungen

Mit der Formulierung eines (hoffentlich) geeigneten statistischen Modells für ein konkretes Zufallsexperiment wie in den vorangegangenen Abschnitten 5.1 und 5.2 ist sicherlich schon einiges gewonnen. Aber was fängt man jetzt mit dem Parameter  $\vartheta$  an, der in das Verteilungsmodell aufgenommen wurde und dessen wahrer Wert *unbekannt* ist?

Das formulierte statistische Modell bildet eine Basis für eine *modellbasierte Datenanalyse*: Aus dem Ergebnis des Zufallsexperiments (den “Beobachtungsdaten”) sollen Rückschlüsse über den wahren Wert des Parameters  $\vartheta$  gezogen werden. Vereinfacht gesagt ist das deshalb möglich, weil das beobachtete Ergebnis  $x$  unter den verschiedenen prinzipiell möglichen Parameterwerten  $\vartheta \in \Theta$  verschiedene Wahrscheinlichkeiten besitzt.

#### Beispiel 5.11 (Binomialmodell)

Betrachte das Binomialmodell von Bsp. 5.3, d. h. eine binomial- $(n, p)$ -verteilte Zufallsvariable  $X$ , wobei  $p \in ]0, 1[$  der Parameter ist (während  $n$  bekannt ist). Angenommen, wir haben einen konkreten Wert  $x \in \{0, 1, \dots, n\}$  als Ergebnis erhalten. Jetzt studieren wir die Wahrscheinlichkeiten, die dieses Ereignis (also das Ergebnis  $x$ ) für alle möglichen Parameterwerte  $p$  besitzt, d. h.

$$P_p(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{für alle } p \in ]0, 1[.$$

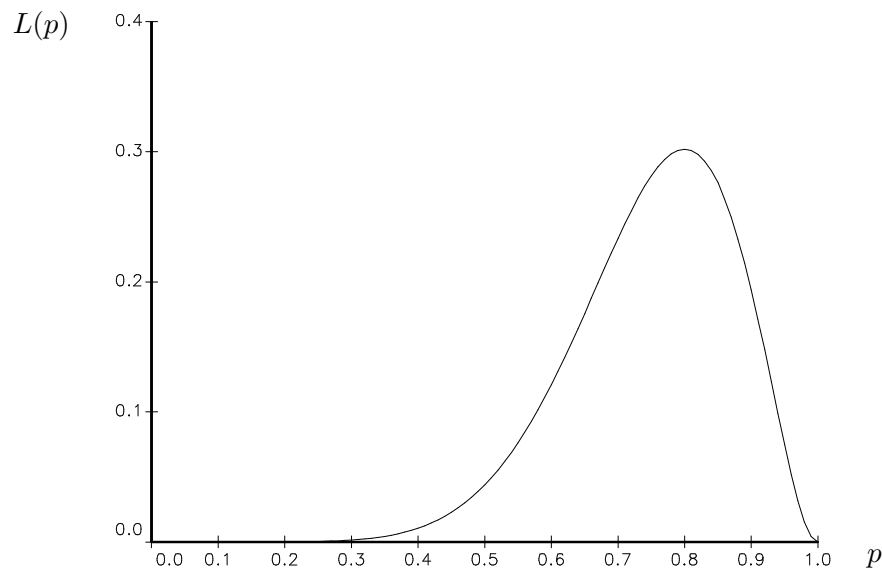


Abb. 5.1: Likelihoodfunktion  $L(p) = P_p(X = 8)$  für eine binomial- $(n = 10, p)$ -verteilte Zufallsvariable  $X$

Wohlgemerkt: Wir betrachten diese Wahrscheinlichkeiten jetzt als Funktion von  $p$  (während  $x$  als der vorliegende Beobachtungswert fest ist). Diese Funktion nennt man auch die *Likelihood-Funktion*. In Abbildung 5.1 ist die Likelihood-Funktion für den Fall  $n = 10$  und  $x = 8$  dargestellt. Die Information, die wir daraus beziehen können, ist in etwa die, dass der wahre Wert von  $p$  in der Nähe von 0.8 liegen sollte. Wesentlich von 0.8 abweichende Werte  $p$  hingegen, z. B.  $p = 0.3$ , erscheinen eher unwahrscheinlich, obwohl natürlich prinzipiell möglich (denn die Wahrscheinlichkeit,  $x = 8$  zu beobachten, gegeben  $p = 0.3$ , ist zwar klein, aber nicht Null). Wenn wir einen Schätzwert für den wahren Wert des Parameters  $p$  angeben wollen, dann bietet sich dafür sicherlich  $\hat{p}(x = 8) = 0.8$  an, da dieser zur maximalen Wahrscheinlichkeit für die vorliegende Beobachtung  $x = 8$  führt. Damit ist eine allgemeine und oft verwendete Schätzmethode angesprochen, die *Maximum Likelihood Methode*, die wir in Kapitel 6 ausführlicher besprechen wollen. ■

Die Beobachtungsdaten (in Beispiel 5.11 ein Wert  $x \in \{0, 1, \dots, n\}$ ) beinhalten eine *statistische* Information über wahren Wert des Parameters  $\vartheta$  (in Beispiel 5.11 heißt dieser Parameter  $p$ ). *Statistische* Information bedeutet hier *unscharfe* Information: Denn die Beobachtungsdaten erlauben in aller Regel keine eindeutige Identifizierung des wahren Parameterwertes  $\vartheta$ . Prinzipiell sind nach wie vor alle Werte  $\vartheta \in \Theta$  möglich, aber im Lichte der Beobachtungsdaten kommen ihnen unterschiedliche Wahrscheinlichkeiten (besser: unterschiedliche “Likelihoods”) zu.

Oben klang schon eine Fragestellung der modellbasierten Datenanalyse an:

### • Konstruktion von Punktschätzungen<sup>1</sup>

d. h.: Auf Grund der Beobachtungsdaten ist eine Schätzung für den wahren Wert des Parameters  $\vartheta$  oder einer Komponente von  $\vartheta$  zu finden. Erwünscht ist natürlich eine möglichst *genaue* Schätzung (diese Zielsetzung zu präzisieren, ist allerdings nicht einfach). Eine wichtige Methode zur Konstruktion von Punktschätzungen, die *Maximum-Likelihood-Methode*, betrachten wir im nachfolgenden Kapitel 6.

Konzepte zur Erfassung der Güte (Genauigkeit) einer Schätzung erfordern weitere Überlegungen, auf die wir hier aus Zeitgründen nicht eingehen können.

<sup>1</sup>Die Bezeichnung Punktschätzung verwenden wir hier zur Abgrenzung vom Begriff Bereichsschätzung oder Intervallschätzung (s. unten). Es soll eben (auf Grund der Beobachtungsdaten) ein “Punkt” des Parameterbereichs  $\Theta$  als Schätzung für  $\vartheta$  gefunden werden.

Zwei weitere Problemkreise der modellbasierten statistischen Datenanalyse können wir hier nur sehr kurz ansprechen:

### • Intervallschätzungen

D.h.: Auf Grund der Beobachtungsdaten ist ein Intervall zu konstruieren, das den wahren Wert des Parameters  $\vartheta$  “mutmaßlich” enthält. Im Fall eines mehrdimensionalen Parameters  $\vartheta$  soll dies für eine (oder alle) einzelne(n) Komponenten von  $\vartheta$  geschehen. Das Konzept des *Konfidenzintervalls* präzisiert “mutmaßlich” dahingehend, dass das Intervall den wahren Parameterwert  $\vartheta$  (bzw. die betrachtete Komponente hiervon) *mit großer Wahrscheinlichkeit* enthalten soll.

#### Beispiel 5.12 (Konfidenzintervall für $\mu$ im Normalvert. modell, vgl. Beispiel 5.8(b))

Im Normalverteilungsmodell mit  $n \geq 2$  u.i.v. normal- $(\mu, \sigma)$ -verteilten Zufallsvariablen von Beispiel 5.8(b) verwendet man als Intervallschätzung für den Parameter  $\mu$ :

$$I(x) = \left[ \bar{x} - \frac{k}{\sqrt{n}} s(x), \bar{x} + \frac{k}{\sqrt{n}} s(x) \right] .$$

Dabei bezeichnen  $x = (x_1, \dots, x_n)$  den Vektor der Beobachtungswerte,  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  (eine Schätzung für  $\mu$ ) und  $s(x) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$  (eine Schätzung für  $\sigma$ ). Der Wert  $k > 0$  ist so zu wählen, dass die Intervallschätzung eine gewünschte *Überdeckungswahrscheinlichkeit*  $1 - \alpha$  für den Parameter  $\mu$  aufweist, wobei  $\alpha$  ein vorgegebener Wert zwischen 0 und 1 ist (in der Regel klein, z.B.  $\alpha = 0.05$ , was einer gewünschten Überdeckungswahrscheinlichkeit von  $1 - \alpha = 0.95$ , also 95% entspricht). Die Präzisierung von *Überdeckungswahrscheinlichkeit* lässt sich mit Hilfe der Zufallsvariablen  $X_1, \dots, X_n$  des Modells unter Einbeziehung aller möglichen Parameterwerte  $\mu$  und  $\sigma$  so fassen:

Für jedes  $\mu$  und jedes  $\sigma > 0$  soll gelten:

$$P_{\mu, \sigma} \left( I(X) \text{ enthält } \mu \right) \geq 1 - \alpha .$$

Hier wurden die konkreten Beobachtungsdaten  $x = (x_1, \dots, x_n)$  durch die Zufallsvariablen  $X = (X_1, \dots, X_n)$  ersetzt, und  $I(X)$  ist daher ein “zufälliges” Intervall, da seine Grenzen die Zufallsvariablen

$$\bar{X} \pm \frac{k}{\sqrt{n}} s(X)$$

sind. Die Philosophie ist dabei folgende: Wenn die W'keit des Ereignisses  $\{I(X) \ni \mu\}$  stets (d.h. egal welche Werte  $\mu$  und  $\sigma$  annehmen) mindestens  $1 - \alpha$  beträgt, dann können wir  $(1 - \alpha) \cdot 100\%$ -iges Vertrauen darin haben, dass für die beobachteten Werte  $x = (x_1, \dots, x_n)$  der Zufallsvariablen  $X = (X_1, \dots, X_n)$  und den wahren Wert  $\mu$  gilt:  $I(x) \ni \mu$ . Man nennt daher auch  $1 - \alpha$  die *Vertrauenswahrscheinlichkeit* des (konkret berechneten) Intervalls  $I(x)$  und nennt dieses ein  $(1 - \alpha)$ -Konfidenzintervall für  $\mu$ .

Die Wahl von  $k > 0$ , die das Intervall  $I(x)$  in diesem Sinne zu einem  $(1 - \alpha)$ -Konfidenzintervall macht, ist:

$$k = t_{n-1, 1-\frac{\alpha}{2}} ,$$

das  $(1 - \frac{\alpha}{2})$ -Quantil der  $t$ -Verteilung mit  $n - 1$  Freiheitsgraden.

Einige solcher Quantile sind in Tafel A.3 des Anhangs angegeben.

### • Lösen von Testproblemen

Aufgrund der Beobachtungsdaten soll entschieden werden, ob der wahre Parameterwert  $\vartheta$  in einer gegebenen (echten) Teilmenge  $\Theta_0 \subset \Theta$  liegt oder nicht (also in  $\Theta_1 = \Theta \setminus \Theta_0$  liegt). Als Kurzschreibweise für diese Fragestellung ist gebräuchlich:

$$H_0 : \vartheta \in \Theta_0 \quad \text{gegen} \quad H_1 : \vartheta \in \Theta_1 ;$$



die mit  $H_0$  bezeichnete Möglichkeit wird *Nullhypothese*, die mit  $H_1$  bezeichnete wird *Alternativhypothese* genannt.

**Beispiel 5.13 (Zweiseitiger  $t$ -Test für  $\mu$  im Normalvert. modell)**

In dem selben Normalverteilungsmodell wie oben (Beispiel 5.12) betrachten wir ein sog. zweiseitiges Testproblem für den Parameter  $\mu$ :

$$H_0 : \mu = \mu_0 \quad \text{gegen} \quad H_1 : \mu \neq \mu_0 ,$$

wobei  $\mu_0$  ein gegebener Wert ist. Z.B. kann diesem Testproblem die Fragestellung unterliegen, ob in einem Produktionsprozess der Sollwert  $\mu_0$  für eine Produktcharakteristik im Mittel eingehalten wird oder nicht. In der allgemeinen Terminologie wäre hier, wenn wie üblich  $\Theta$  die Menge aller möglichen Paare  $\vartheta = (\mu, \sigma)$  bezeichnet,

$$\Theta_0 = \{ \vartheta = (\mu, \sigma) \in \Theta : \mu = \mu_0 \} \quad \text{und} \quad \Theta_1 = \{ \vartheta = (\mu, \sigma) \in \Theta : \mu \neq \mu_0 \} .$$

Für dieses Testproblem verwendet man den zweiseitigen  $t$ -Test:

$$\varphi(x) = \begin{cases} 1 \\ 0 \end{cases} , \text{ falls } \sqrt{n} \frac{|\bar{x} - \mu_0|}{s(x)} \begin{matrix} > \\ \leq \end{matrix} c ,$$

wobei wieder  $x = (x_1, \dots, x_n)$  den Vektor der Beobachtungswerte,  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  und  $s(x) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$  bezeichnen; die Werte 1 und 0 für die Entscheidungsregel ("Test")  $\varphi(x)$  stehen dabei für *Entscheidung für  $H_1$*  bzw. *Entscheidung für  $H_0$* . Der *kritische Wert*  $c$  ist geeignet zu wählen; üblicherweise orientiert man sich dabei an der *Fehlerwahrscheinlichkeit erster Art* des Tests  $\varphi$ , die man durch geeignete Wahl von  $c$  durch einen gewünschten (kleinen) Wert  $\alpha$  (z.B.  $\alpha = 0.05$ ) nach oben beschränken kann. Der *Fehler erster Art* wird (bei Verwenden der Entscheidungsregel  $\varphi$ ) begangen, wenn in Wirklichkeit die Nullhypothese  $H_0$  zutrifft, aber auf Grund der Beobachtung  $x$  herauskommt  $\varphi(x) = 1$ , also für  $H_1$  entschieden wird. Es stellt sich heraus, dass für  $c$  (zu einem gegebenen Wert  $\alpha \in ]0, 1[$ ) wieder zu nehmen ist (vergl. Beispiel 5.12):

$$c = t_{n-1, 1-\frac{\alpha}{2}} ,$$

das  $(1 - \frac{\alpha}{2})$ -Quantil der  $t$ -Verteilung mit  $n - 1$  Freiheitsgraden. Man nennt  $\varphi$  den zweiseitigen  $t$ -test zum Signifikanzniveau  $\alpha$ .

Der Begriff der *Fehlerwahrscheinlichkeit(en) erster Art* lässt sich mit Hilfe der Zufallsvariablen  $X_1, \dots, X_n$  des Modells so präzisieren:

$$P_{\mu_0, \sigma}(\varphi(X) = 1) \quad \text{für alle } \sigma > 0 .$$

Diese sollen die gewünschte Schranke  $\alpha$  nicht überschreiten. d.h.

$$P_{\mu_0, \sigma}(\varphi(X) = 1) \leq \alpha \quad \text{für alle } \sigma > 0 ,$$

und dies wird durch die genannte Wahl des kritischen Wertes  $c$  gewährleistet.

**Beispiel 5.14 ( $t$ -Test im 2-Stichproben-Normalvert.modell, vergl. Beisp. 5.9)**

Im 2-Stichproben-Normalverteilungsmodell von Beispiel 5.9 betrachten wir das Testproblem über die Parameter  $\mu_1$  und  $\mu_2$  der beiden zu Grunde liegenden Normalverteilungen,

$$H_0 : \mu_1 \leq \mu_2 \quad \text{gegen} \quad H_1 : \mu_1 > \mu_2 .$$

Dieses Testproblem drückt z.B. die Fragestellung aus, ob ein alternatives Herstellungsverfahren Bauteile mit einer höheren mittleren Festigkeit liefert als ein herkömmliches Verfahren. Bezeichne  $x = (x_1, \dots, x_{n_1})$  die Beobachtungswerte der ersten Stichprobe (z.B. Festigkeiten unter Verfahren

1) und  $y = (y_1, \dots, y_{n_2})$  die Beobachtungswerte der zweiten Stichprobe (z.B. Festigkeiten unter Verfahren 2). Der *Zwei-Stichproben  $t$ -Test* für das genannte Testproblem lautet dann:

$$\varphi(x, y) = \begin{cases} 1 \\ 0 \end{cases}, \text{ falls } \begin{matrix} T(x) > \\ \leq \end{matrix} c,$$

mit der *Teststatistik*  $T(x, y)$  :

$$T(x, y) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\bar{y} - \bar{x}}{s(x, y)}, \quad \text{wobei}$$

$$\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i, \quad \bar{y} = \frac{1}{n_2} \sum_{j=1}^{n_2} y_j,$$

$$s(x, y) = \sqrt{\frac{1}{n_1 + n_2 - 2} \left[ \sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{j=1}^{n_2} (y_j - \bar{y})^2 \right]}.$$

Der kritische Wert  $c$  wird wieder in Abhängigkeit eines gegebenen  $\alpha \in ]0, 1[$  so gewählt, dass die Fehlerwahrscheinlichkeiten erster Art des Tests  $\varphi$  höchstens  $\alpha$  betragen, und dies führt zur Wahl

$$c = t_{n_1 + n_2 - 2, 1 - \alpha},$$

das  $(1 - \alpha)$ -Quantil der  $t$ -Verteilung mit  $n_1 + n_2 - 2$  Freiheitsgraden. ■

Die genannten drei Problemkreise (Punktschätzung, Intervallschätzung, Test) greifen sicherlich ineinander, wie die obigen drei Beispiele auch erkennen lassen. So wird man eine Intervallschätzung mit Hilfe einer Punktschätzung gewinnen können, indem um diese herum eine gewisse Toleranz zugelassen wird, welche die Unsicherheit der Punktschätzung berücksichtigt. Zur Lösung eines Testproblems wird man ebenfalls eine Punkt- oder Intervallschätzung benötigen, um zu einer fundierten Entscheidung zwischen  $H_0$  und  $H_1$  zu kommen. Im folgenden Kapitel betrachten wir eine populäre Methode zur Konstruktion von Punktschätzungen.

## Kapitel 6

# Maximum-Likelihood-Schätzung

### 6.1 Likelihood-Funktion

In Beispiel 5.11 hatten wir schon in einem speziellen Fall den Begriff der *Likelihood-Funktion* verwendet. Dieser soll jetzt allgemeiner gefasst werden, bezogen auf möglichst viele statistische Modelle. Mit den uns zur Verfügung stehenden w-theoretischen Grundlagen können wir dies leisten zum einen für *diskrete* Modelle (d.h. die Zufallsvariablen des Modells sind *diskrete* Zufallsvariablen) und zum anderen für *absolut-stetige* Modelle (d.h. die Zufallsvariablen des Modells sind reelle, *absolut-stetig* verteilte Zufallsvariablen). Wenn das Modell *mehrere* Zufallsvariablen aufweist, dann werden wir diese als stochastisch unabhängig voraussetzen. Das entspricht der häufigen Situation, dass ein Zufallsexperiment aus mehreren *unabhängigen* Telexperimenten besteht. Beginnen wir mit diskreten statistischen Modellen.

#### Definition 6.1 (Diskretes Modell mit einer Zufallsvariablen)

Sei  $X$  eine diskrete Zufallsvariable, d.h.  $X : \Omega \longrightarrow M$  mit einer abzählbaren Menge  $M$ , und die Verteilung von  $X$  beinhalte einen unbekannten (möglicherweise mehrdimensionalen) Parameter  $\vartheta \in \Theta$ , d.h. wir haben die Wahrscheinlichkeiten

$$P_{\vartheta}(X = x) \quad \text{für alle } x \in M \text{ und alle } \vartheta \in \Theta.$$

Lesen wir diese Wahrscheinlichkeiten als Funktion von  $\vartheta \in \Theta$  für ein festes  $x \in M$  (das als beobachteter Wert der Zufallsvariablen  $X$  interpretiert wird), dann haben wir die Likelihood-Funktion (zur Beobachtung  $x$ ), d.h.

$$L(\vartheta) = P_{\vartheta}(X = x) \quad \text{für alle } \vartheta \in \Theta.$$

**Definition 6.2 (Diskretes Modell mit mehreren stoch. unabh. ZV'en)**

Seien  $X_1, X_2, \dots, X_n$  stochastisch unabhängige diskrete Zufallsvariablen,

$$X_i : \Omega \longrightarrow M_i, \quad i = 1, 2, \dots, n,$$

mit abzählbaren Mengen  $M_1, M_2, \dots, M_n$ . Die Verteilung jeder Zufallsvariablen  $X_i$  beinhalte einen unbekannten (möglicherweise mehrdimensionalen) Parameter  $\vartheta \in \Theta$ . Wir haben daher Produkt-Wahrscheinlichkeiten

$$P_{\vartheta}(X_1 = x_1) \cdot P_{\vartheta}(X_2 = x_2) \cdot \dots \cdot P_{\vartheta}(X_n = x_n)$$

für alle  $x_1 \in M_1, x_2 \in M_2, \dots, x_n \in M_n$  und alle  $\vartheta \in \Theta$ . Lesen wir diese Produkt-Wahrscheinlichkeiten als Funktion von  $\vartheta \in \Theta$  für feste  $x_1 \in M_1, x_2 \in M_2, \dots, x_n \in M_n$  (die als beobachtete Werte der Zufallsvariablen  $X_1, X_2, \dots, X_n$  interpretiert werden), dann haben wir die Likelihood-Funktion (zu den Beobachtungen  $x_1, x_2, \dots, x_n$ ), d. h.

$$L(\vartheta) = P_{\vartheta}(X_1 = x_1) \cdot P_{\vartheta}(X_2 = x_2) \cdot \dots \cdot P_{\vartheta}(X_n = x_n)$$

Kommen wir nun zu statistischen Modellen mit einer (oder mehreren) *stetig-verteilten* reellen Zufallsvariablen. Beginnen wir mit dem Fall einer Zufallsvariablen.

Gegeben sei also ein Modell mit einer reellen Zufallsvariablen  $X : \Omega \longrightarrow \mathbb{R}$ , deren Verteilung einen (evtl. mehr-dimensionalen) Parameter  $\vartheta \in \Theta$  beinhaltet. Die Zufallsvariable sei stetig-verteilt für alle  $\vartheta$ , d. h. für jedes  $\vartheta \in \Theta$  ist ihre Verteilungsfunktion

$$F(x; \vartheta) = P_{\vartheta}(X \leq x), \quad x \in \mathbb{R}$$

eine stetige Funktion (von  $x$ ). In Abschnitt 3.3 haben wir gesehen, dass die Stetigkeit der Verteilungsfunktionen gleichbedeutend ist mit:

$$P_{\vartheta}(X = x) = 0 \quad \text{für alle } x \in \mathbb{R} \text{ und alle } \vartheta \in \Theta.$$

Was also sollten wir hier unter der *Likelihood-Funktion* verstehen?

Eine direkte Übertragung von Definition 6.1 auf die vorliegende Situation wäre sinnlos, da dies als Likelihood-Funktion die Konstante Null liefern würde. Ein Ausweg mag sein, dass wir den beobachteten Wert  $x$  der Zufallsvariablen  $X$  etwas "aufweichen": Statt des Ereignisses  $\{X = x\}$  (und den ihm für alle  $\vartheta \in \Theta$  zukommenden Wahrscheinlichkeiten, die alle gleich 0 sind) betrachten wir für irgendeine (kleine) positive Zahl  $\varepsilon$  das Ereignis

$$\{x - \varepsilon \leq X \leq x + \varepsilon\},$$

welches ja auch eintritt, wenn  $\{X = x\}$  eintritt, und wir betrachten die Wahrscheinlichkeiten dieses Ereignisses:

$$P_{\vartheta}(x - \varepsilon \leq X \leq x + \varepsilon) = F(x + \varepsilon; \vartheta) - F(x - \varepsilon; \vartheta),$$

(vgl. Abschnitt 3.3). Diese Wahrscheinlichkeiten sind in der Regel positiv und variieren mit  $\vartheta \in \Theta$ . Jedoch hängen sie noch von dem willkürlich gewählten Wert  $\varepsilon > 0$  ab. Auch ist die Intention, den Wert  $\varepsilon > 0$  sehr klein zu wählen, damit die Information  $\{X = x\}$  noch hinreichend scharf durch das Ereignis  $\{x - \varepsilon \leq X \leq x + \varepsilon\}$  erfasst wird. Für  $\varepsilon \rightarrow 0$  konvergieren die Wahrscheinlichkeiten dieses Ereignisses wiederum gegen Null. Daher erscheint es sinnvoll, die Wahrscheinlichkeiten ins Verhältnis zur Intervalllänge  $2\varepsilon$  zu setzen und den Grenzwert *des Quotienten* für  $\varepsilon \rightarrow 0$  zu betrachten:

$$\lim_{\varepsilon \rightarrow 0} \frac{P_{\vartheta}(x - \varepsilon \leq X \leq x + \varepsilon)}{2\varepsilon} = \lim_{\varepsilon \rightarrow 0} \frac{F(x + \varepsilon; \vartheta) - F(x - \varepsilon; \vartheta)}{2\varepsilon}.$$

Wenn wir nun voraussetzen, dass die Zufallsvariable  $X$  absolut-stetig verteilt ist (s. Definition 3.4), also für jedes  $\vartheta \in \Theta$  eine Dichtefunktion  $f(x; \vartheta)$  (nämlich die Ableitung der Verteilungsfunktion  $F(x; \vartheta)$  nach  $x$ ) besitzt, dann erhalten wir für den obigen Grenzwert:

$$\begin{aligned} \frac{F(x + \varepsilon; \vartheta) - F(x - \varepsilon; \vartheta)}{2\varepsilon} &= \frac{1}{2} \frac{F(x + \varepsilon; \vartheta) - F(x; \vartheta)}{\varepsilon} + \frac{1}{2} \frac{F(x; \vartheta) - F(x - \varepsilon; \vartheta)}{\varepsilon} \\ &\xrightarrow{(\varepsilon \rightarrow 0)} \frac{1}{2} f(x; \vartheta) + \frac{1}{2} f(x; \vartheta) = f(x; \vartheta). \end{aligned}$$

(Man erinnere sich an die Definition der Ableitung einer reellen Funktion als Grenzwert von *Differenzenquotienten*.) Damit sind wir für den Fall einer absolut-stetig verteilten reellen Zufallsvariablen zu folgender Definition der Likelihood-Funktion geführt.

**Definition 6.3 (Modell mit einer absolut-stetig verteilten ZV'en)**

Sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable, deren Verteilung einen Parameter  $\vartheta \in \Theta$  beinhaltet und die für jedes  $\vartheta \in \Theta$  eine Dichtefunktion  $f(x; \vartheta)$ ,  $x \in \mathbb{R}$ , besitzt.

Lesen wir die Werte der Dichtefunktionen als Funktion von  $\vartheta \in \Theta$  für ein festes  $x \in \mathbb{R}$  (das als beobachteter Wert der Zufallsvariablen  $X$  interpretiert wird), dann haben wir die Likelihood-Funktion (zur Beobachtung  $x$ ), d. h.

$$L(\vartheta) = f(x; \vartheta) \quad \text{für alle } \vartheta \in \Theta.$$

Für ein statistisches Modell mit  $n \geq 2$  absolut-stetig verteilten und stochastisch unabhängigen reellen Zufallsvariablen  $X_1, X_2, \dots, X_n$  können wir ähnlich (aber etwas komplizierter) argumentieren: Seien die (reellen) Beobachtungswerte  $x_1, x_2, \dots, x_n$  gegeben. Als Likelihood nehmen wir

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{(2\varepsilon)^n} P_\vartheta(x_1 - \varepsilon \leq X_1 \leq x_1 + \varepsilon, \dots, x_n - \varepsilon \leq X_n \leq x_n + \varepsilon).$$

Wegen der stochastischen Unabhängigkeit von  $X_1, \dots, X_n$  ist

$$\begin{aligned} &\frac{1}{(2\varepsilon)^n} P_\vartheta(x_1 - \varepsilon \leq X_1 \leq x_1 + \varepsilon, \dots, x_n - \varepsilon \leq X_n \leq x_n + \varepsilon) \\ &= \prod_{i=1}^n \left[ \frac{1}{2\varepsilon} P_\vartheta(x_i - \varepsilon \leq X_i \leq x_i + \varepsilon) \right], \end{aligned}$$

und für  $\varepsilon \rightarrow 0$  konvergiert dies gegen

$$\prod_{i=1}^n f_{X_i}(x_i; \vartheta),$$

die Produkt-Dichtefunktion. Wir gelangen daher zu folgender Definition.

**Definition 6.4 (Modell mit mehreren stoch. unabh. abs. stetig-verteilten ZV'en)**

Seien  $X_1, X_2, \dots, X_n$  reelle Zufallsvariablen mit Dichtefunktionen  $f_{X_i}(x_i; \vartheta)$  ( $i = 1, \dots, n$ ), wobei  $\vartheta \in \Theta$  der Parameter (evtl. mehrdimensional) des Modells ist. Lesen wir die Produkt-Dichtefunktion als Funktion von  $\vartheta \in \Theta$  für festen  $x_1, x_2, \dots, x_n \in \mathbb{R}$  (die als beobachtete Werte der Zufallsvariablen  $X_1, X_2, \dots, X_n$  interpretiert werden), dann haben wir die Likelihood-Funktion (zu den Beobachtungen  $x_1, x_2, \dots, x_n$ ),

$$L(\vartheta) = f_{X_1}(x_1; \vartheta) \cdot f_{X_2}(x_2; \vartheta) \cdot \dots \cdot f_{X_n}(x_n; \vartheta) \quad \text{für alle } \vartheta \in \Theta.$$

**Bemerkung 6.5 (Log-Likelihood-Funktion)**

Da Wahrscheinlichkeiten und Dichtefunktionen stets nicht-negativ sind, gilt für eine Likelihood-Funktion gemäß Definitionen 6.1, 6.2, 6.3 und 6.4 stets

$$L(\vartheta) \geq 0 \quad \text{für alle } \vartheta \in \Theta.$$

Wenn die Likelihood-Funktion strikt positiv ist (also  $L(\vartheta) = 0$  nicht auftritt), dann können wir die *Log-Likelihood-Funktion* bilden,

$$\ell(\vartheta) = \ln L(\vartheta) \quad \text{für alle } \vartheta \in \Theta.$$

Da der Logarithmus eine streng monoton wachsende Funktion ist, bleiben beim Übergang von der Likelihood-Funktion zur Log-Likelihood-Funktion die Ordnungsrelationen zwischen Funktionswerten verschiedener  $\vartheta$ -Werte erhalten, d. h. es gilt für je zwei Werte  $\vartheta, \vartheta' \in \Theta$ :

$$L(\vartheta) \leq L(\vartheta') \quad \text{genau dann, wenn} \quad \ell(\vartheta) \leq \ell(\vartheta').$$

Insbesondere haben die beiden Probleme

$$L(\vartheta) \longrightarrow \max_{\vartheta \in \Theta}! \quad \text{und} \quad \ell(\vartheta) \longrightarrow \max_{\vartheta \in \Theta}!$$

dieselben Optimallösungen für  $\vartheta$ .

Die Log-Likelihood-Funktion hat oftmals eine einfachere Gestalt als die Likelihood-Funktion (z. B. werden Produkte zu Summen, allerdings um den Preis, dass jetzt Logarithmen auftreten).

Ein Beispiel für eine Likelihood-Funktion in einem diskreten Modell wurde schon in Beispiel 5.11 (eine binomial- $(n, p)$ -verteilte Zufallsvariable) gegeben:

$$L(p) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{für alle } p \in ]0, 1[ ,$$

und die Log-Likelihood-Funktion lautet hier

$$\ell(p) = \ln \binom{n}{x} + x \ln(p) + (n-x) \ln(1-p).$$

Ein Beispiel für ein Modell mit einer absolut-stetig verteilten Zufallsvariablen:

Im Modell mit einer exponential- $(\lambda)$ -verteilten Zufallsvariablen  $X$  (vgl. Beispiel 5.7(a)) erhalten wir gemäß Definition 6.3 als Likelihood-Funktion zu einer Beobachtung  $x > 0$ :

$$L(\lambda) = f(x; \lambda) = \lambda \exp(-\lambda x) \quad \text{für alle } \lambda > 0,$$

und als Log-Likelihood-Funktion

$$\ell(\lambda) = \ln(\lambda) - \lambda x \quad \text{für alle } \lambda > 0.$$

Wir haben die Beobachtung  $x$  als positiv angenommen, da eine exponential-verteilte Zufallsvariable mit Wahrscheinlichkeit 1 einen positiven Wert liefert. Formal würden wir für eine Beobachtung  $x \leq 0$  als Likelihood-Funktion die Konstante 0 erhalten. ■

## 6.2 Maximum-Likelihood-Schätzung

Eine Methode zur Konstruktion einer Punktschätzung für den unbekannten Parameter  $\vartheta$  in einem statistischen Modell auf Grund der Beobachtungsdaten  $(x$  oder  $x_1, x_2, \dots, x_n)$  ist die *Maximum-Likelihood-Methode* (Abk.: *ML-Methode*). Diese ist sehr plausibel: Unter den vorliegenden Beobachtungsdaten wird derjenige Parameterwert als Schätzwert für  $\vartheta$  genommen, der den größten Likelihood-Wert besitzt.

Um die Notation zu vereinfachen, werden wir oft von der Beobachtung  $x$  sprechen, auch wenn diese eigentlich mehrdimensional ist und aus mehreren Einzelbeobachtungen  $x_1, \dots, x_n$  besteht, also  $x = (x_1, \dots, x_n)$ . Aus dem Zusammenhang (Modell mit einer oder mehreren Zufallsvariablen) sollte stets klar sein, ob  $x$  ein- oder mehrdimensional ist. Ebenso werden wir oft von dem Parameter  $\vartheta$  sprechen, auch wenn dieser mehrdimensional ist.

### Definition 6.6 (ML-Schätzung)

Gegeben sei ein statistisches Modell mit dem Parameter  $\vartheta \in \Theta$ . Zur Beobachtung  $x$  sei  $L(\vartheta)$ ,  $\vartheta \in \Theta$ , die Likelihood-Funktion. Dann heißt ein Wert

$$\hat{\vartheta} = \hat{\vartheta}(x) \in \Theta$$

eine Maximum-Likelihood-Schätzung (ML-Schätzung) für  $\vartheta$ , wenn gilt:

$$L(\hat{\vartheta}) = \max_{\vartheta \in \Theta} L(\vartheta) .$$

Wir werden im Folgenden eine ML-Schätzung für  $\vartheta$  auf Grund der Beobachtung  $x$  stets mit  $\hat{\vartheta}(x)$  bezeichnen (also die Abhängigkeit von  $x$  explizit anführen), um auch in der Notation auszudrücken, dass es sich hierbei um einen konkreten Schätzwert auf Grund der konkreten Beobachtung  $x$  handelt.

Die ML-Methode zur Parameter(-punkt-)schätzung erfreut sich breiter Anwendung in der modellbasierten Datenanalyse. Ein Grund dafür ist, dass sie in den einfachsten (aber standardmäßig verwendeten) Modellen in der Tat sehr plausible Schätzungen liefert (s. nachfolgende Beispiele). Ein weiterer Grund ist, dass sich ML-Schätzungen in der Regel als "gute" Schätzungen mathematisch nachweisen lassen, worauf wir hier aber nicht eingehen können.

### Beispiel 6.7

#### (a) u.i.v. 0-1-Variablen. (vgl. Beispiel 5.1(a))

Betrachte das Modell mit  $n$  u.i.v. 0-1-wertigen Zufallsvariablen  $X_1, X_2, \dots, X_n$ ,

$$P_p(X_i = 1) = p \quad \text{für alle } i = 1, 2, \dots, n \text{ und alle } p \in ]0, 1[ .$$

Der Parameter ist  $p \in ]0, 1[$ . Seien  $x_1, x_2, \dots, x_n \in \{0, 1\}$  die Beobachtungen. Wir berechnen die Likelihood-Funktion. Gemäß Definition 6.2 ist

$$L(p) = P_p(X_1 = x_1) P_p(X_2 = x_2) \cdots P_p(X_n = x_n) .$$

Nun sind auf der rechten Seite die Faktoren jeweils entweder gleich  $p$  (wenn  $x_i = 1$ ) oder gleich  $1 - p$  (wenn  $x_i = 0$ ). Die Anzahl der "1"en unter den Beobachtungen  $x_1, x_2, \dots, x_n$  können wir als  $\sum_{i=1}^n x_i$  schreiben und die Anzahl der "0"en als  $n - \sum_{i=1}^n x_i$ . Wir erhalten:

$$L(p) = p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i} ,$$

und für die Log-Likelihood-Funktion

$$\ell(p) = \left( \sum_{i=1}^n x_i \right) \ln(p) + \left( n - \sum_{i=1}^n x_i \right) \ln(1 - p) .$$

Eine "Kurvendiskussion" der Funktion  $\ell(p)$  für  $p \in ]0, 1[$  liefert als Maximalstelle

$$\hat{p}(x) = \frac{\sum_{i=1}^n x_i}{n} ,$$

also eine sehr plausible Schätzung für  $p$  (die relative Häufigkeit als Schätzung für die “theoretische” Wahrscheinlichkeit  $p$ ).

Streng genommen ist die ML-Schätzung  $\hat{p}(x)$  in den beiden extremen Datensituationen, dass nämlich  $\sum_{i=1}^n x_i = 0$  ist (also alle  $x_i = 0$  sind) oder  $\sum_{i=1}^n x_i = n$  ist (also alle  $x_i = 1$  sind), nicht zulässig als Schätzung für  $p \in ]0, 1[$ . In diesen Fällen ist ja  $\hat{p}(x) = 0$  bzw.  $\hat{p}(x) = 1$ , und diese Werte gehören nicht mehr zum Parameterbereich  $]0, 1[$ . Ähnliche Ausnahmen werden auch für andere Modelle auftreten. Diese sind aber vernachlässigbar oder nahezu vernachlässigbar.

Auf die “Kurvendiskussion” der Log-Likelihood-Funktion, die zur genannten Formel für die ML-Schätzung  $\hat{p}(x)$  führt, sind wir nicht weiter eingegangen. Der wichtigste Schritt ist dabei die Berechnung der Ableitung  $\ell'(p)$  und die Lösung der Gleichung  $\ell'(p) = 0$ . Man erhält:

$$\ell'(p) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p},$$

und Multiplikation beider Seiten der Gleichung  $\ell'(p) = 0$  mit  $p(1-p)$  liefert

$$\left(\sum_{i=1}^n x_i\right)(1-p) - \left(n - \sum_{i=1}^n x_i\right)p = 0,$$

d. h.

$$\left(\sum_{i=1}^n x_i\right) - np = 0, \quad \text{also} \quad p = \frac{\sum_{i=1}^n x_i}{n}.$$

**(b) Binomialmodell.** (vgl. Beispiele 5.1(b) und 5.3)

Betrachte das Modell mit einer binomial- $(n, p)$ -verteilten Zufallsvariablen  $X$ , wobei  $p \in ]0, 1[$  der Parameter ist. Die Log-Likelihood-Funktion hatten wir schon in Bemerkung 6.5 formuliert:

$$\ell(p) = \ln \binom{n}{x} + x \ln(p) + (n-x) \ln(1-p),$$

wobei  $x \in \{0, 1, \dots, n\}$  die gegebene Beobachtung ist. Die Log-Likelihood-Funktion ist ähnlich jener im Teil (a) oben: Hier steht  $x$  statt  $\sum_{i=1}^n x_i$ , und es tritt der zusätzliche additive Term  $\ln \binom{n}{x}$  auf, der hier aber nur die Rolle einer Konstanten und damit auf die Maximallösung keinen Einfluß hat. Daher ist klar: Die ML-Schätzung lautet

$$\hat{p}(x) = \frac{x}{n}$$

(wobei streng genommen wiederum die extremen Beobachtungswerte  $x = 0$  und  $x = n$  ausgeschlossen werden müssen).

**Beispiel 6.8 (Hypergeometrisches Modell, vgl. Beispiel 5.4)**

Betrachte das Modell mit einer hypergeometrisch- $(N, s, n)$ -verteilten Zufallsvariablen  $X$ , wobei  $s \in \{0, 1, \dots, N\}$  der Parameter ist. Die Werte  $N \in \mathbb{N}$  und  $n \in \mathbb{N}$  (mit  $n < N$ ) sind bekannt. Zu einer gegebenen Beobachtung  $x \in \{0, 1, \dots, n\}$  lautet die Likelihood-Funktion:

$$L(s) = \binom{s}{x} \binom{N-s}{n-x} / \binom{N}{n}.$$

Der Binomialkoeffizient  $\binom{p}{q}$  ( $p, q \in \mathbb{N}_0$ ) ist nur für  $p \geq q$  positiv, ansonsten ist er gleich 0. Daher ist  $L(s)$  genau dann positiv, wenn  $s \geq x$  und  $N-s \geq n-x$ , d. h. wenn gilt

$$x \leq s \leq N - (n-x).$$



Wir können uns also im Folgenden auf diesen Bereich für  $s$  beschränken, denn nur diese Werte kommen für die ML-Schätzung in Frage (für die übrigen ist ja  $L(s) = 0$ ).

Die Likelihood-Funktion  $L(s)$  hat nun den folgenden Verlauf:

Im Bereich  $x \leq s \leq \left\lfloor \frac{(N+1)x}{n} \right\rfloor$  ist  $L(s)$  monoton wachsend und im Bereich  $\left\lfloor \frac{(N+1)x}{n} \right\rfloor \leq s \leq N - (n - x)$  monoton fallend (dabei bezeichnet  $\lfloor a \rfloor$  die größte ganze Zahl  $\leq a$ ).

Wir erhalten daher als ML-Schätzung:

$$\hat{s}(x) = \begin{cases} \left\lfloor \frac{(N+1)x}{n} \right\rfloor & , \text{ falls } x < n \\ N & , \text{ falls } x = n. \end{cases}$$

Das ist auch eine plausible Schätzung, denn  $\hat{s}(x)$  ist eine der benachbarten ganzen Zahlen des Wertes  $Nx/n$ , also eine ganzzahlige Rundung dieses Wertes, und somit ist

$$\frac{\hat{s}(x)}{N} \approx \frac{x}{n}.$$

Im Kontext von Beispiel 5.4 können wir das so interpretieren: Die Schätzung der Anzahl der “schlechten” Objekte in der Grundgesamtheit erfolgt so, dass der geschätzte relative “Schlechtanteil” in der Grundgesamtheit möglichst nah dem “Schlechtanteil” in der Stichprobe ( $\frac{x}{n}$ ) liegt.

Zahlenbeispiel:  $N = 30$ ,  $n = 7$ ,  $x = 2$  liefert die ML-Schätzung

$$\hat{s}(x) = \left\lfloor \frac{31 \cdot 2}{7} \right\rfloor = \left\lfloor \frac{62}{7} \right\rfloor = \lfloor 8.857 \rfloor = 8 ;$$

die Likelihood-Funktion ist für diesen Fall in Abbildung 7.1 dargestellt.

Um das beschriebene Verhalten der Likelihood-Funktion im allgemeinen Fall zu erkennen, betrachte man die folgenden Quotienten (für alle  $s$  mit  $x \leq s < N - (n - x)$ ):

$$\begin{aligned} \frac{L(s+1)}{L(s)} &= \frac{\binom{s+1}{x} \binom{N-s-1}{n-x}}{\binom{s}{x} \binom{N-s}{n-x}} = \frac{(s+1)! (N-s-1)! x! (s-x)! (n-x)! (N-s-n+x)!}{x! (s+1-x)! (n-x)! (N-s-1-n+x)! s! (N-s)!} \\ &= \frac{(s+1) (N-s-n+x)}{(s+1-x) (N-s)}. \end{aligned}$$

Die Ungleichung

$$\frac{(s+1) (N-s-n+x)}{(s+1-x) (N-s)} > 1$$

lässt sich äquivalent schreiben als

$$(s+1) (N-s-n+x) > (s+1-x) (N-s),$$

und elementare Umformung hiervon liefert

$$s < \frac{(N+1)x}{n} - 1.$$

Analog erhält man für umgekehrte Ungleichung:

$$\frac{(s+1) (N-s-n+x)}{(s+1-x) (N-s)} < 1 \iff s > \frac{(N+1)x}{n} - 1.$$

Für  $x \leq s < (N+1)x/n - 1$  ist also  $L(s) < L(s+1)$ , und für  $(N+1)x/n - 1 < s < N - (n - x)$  ist  $L(s) > L(s+1)$ . Daraus ersieht man das oben beschriebene Verhalten der Likelihood-Funktion.

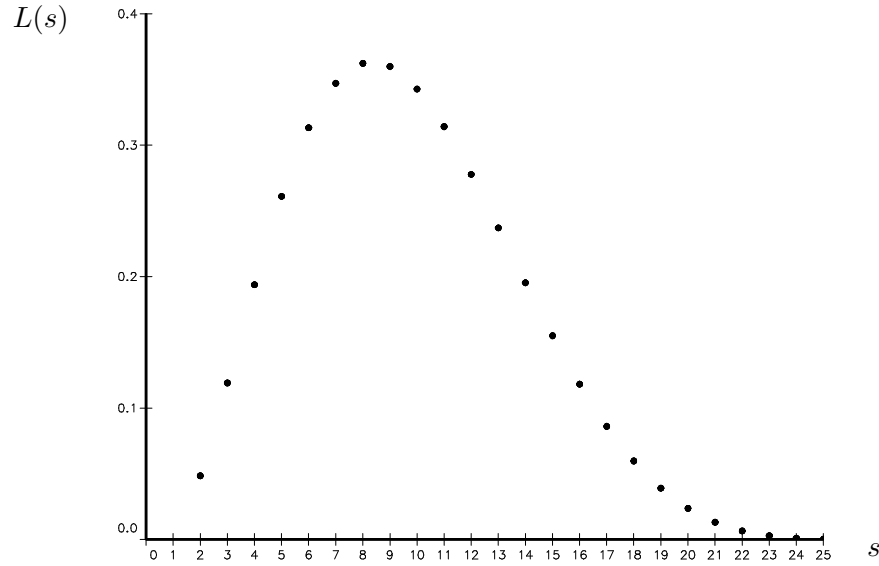


Abb. 6.1: Likelihood-Funktion  $L(s)$  für eine hypergeometrisch- $(N = 30, s, n = 7)$ -verteilte Zufallsvariable  $X$  auf Grund der Beobachtung  $x = 2$

**Beispiel 6.9 (Normalverteilungsmodell, vgl. Beispiel 5.8(b) )**

Betrachte das Modell mit  $n \geq 2$  u.i.v. normal- $(\mu, \sigma)$ -verteilten Zufallsvariablen  $X_1, X_2, \dots, X_n$ . Der Parameter besteht hier aus  $\mu \in \mathbb{R}$  und  $\sigma > 0$ , ist also zwei-dimensional ( $\vartheta = (\mu, \sigma)$ ). Aus Definition 6.4 und Beispiel 5.8(b) erhalten wir für die die Log-Likelihood-Funktion:

$$\ell(\mu, \sigma) = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 .$$

Hinsichtlich der Maximierung von  $\ell(\mu, \sigma)$  erkennen wir sofort: Der Wert für  $\mu$  muss so gewählt werden, dass

$$\sum_{i=1}^n (x_i - \mu)^2 \quad \text{minimal wird.}$$

Das ist leicht zu lösen, da dieser Ausdruck als Funktion von  $\mu$  (für feste Beobachtungsdaten  $x_1, x_2, \dots, x_n$ ) eine quadratische Funktion ist:

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i^2 - 2x_i\mu + \mu^2) = \sum_{i=1}^n x_i^2 - 2 \left( \sum_{i=1}^n x_i \right) \mu + n\mu^2 ,$$

und diese wird minimal, wenn

$$2 \sum_{i=1}^n x_i - 2n\mu = 0 , \quad \text{d.h.} \quad \mu = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} .$$

Als ML-Schätzung für den Parameter  $\mu$  haben wir daher  $\hat{\mu}(x) = \bar{x}$ .

Nach Ersetzen des Parameters  $\mu$  in der Log-Likelihood-Funktion durch seine ML-Schätzung  $\hat{\mu}(x)$  verbleibt die Maximierung bzgl.  $\sigma$  dieser Funktion:

$$\ell(\hat{\mu}(x), \sigma) = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 .$$

Hier ist also eine “Kurvendiskussion” durchzuführen; wir zeigen hier nur den Schritt der Berechnung der Nullstelle der Ableitung:

Die Ableitung nach  $\sigma$  ist

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \bar{x})^2 .$$

Setzen wir diese gleich Null und multiplizieren beide Seiten der Gleichung mit  $\sigma^3$ , so erhalten wir

$$-n\sigma^2 + \sum_{i=1}^n (x_i - \bar{x})^2 = 0 , \quad \text{d.h.} \quad \sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} .$$

Insgesamt haben wir also die (plausiblen) ML-Schätzungen:

$$\hat{\mu}(x) = \bar{x} , \quad \hat{\sigma}(x) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} .$$

Im Fall, dass die Beobachtungswerte  $x_1, x_2, \dots, x_n$  alle gleich sind, also  $x_1 = x_2 = \dots = x_n$  gilt, erhalten wir  $\hat{\sigma}(x) = 0$ . Das ist streng genommen kein zulässiger Schätzwert für  $\sigma > 0$ . Nun ist diese Datensituation sicherlich wenig wahrscheinlich, so dass wir über diesen Fall hinwegsehen können. In der Tat lässt sich beweisen, dass dieser Fall Wahrscheinlichkeit Null hat (also praktisch nicht auftritt, sofern das angenommene Modell zutrifft):  $P_{\mu, \sigma}(X_1 = X_2 = \dots = X_n) = 0$  für alle  $\mu, \sigma$ .

**Beispiel 6.10 (Regressionsgerade/Normalverteilungsmodell, vgl. Beispiel 5.10)**

Wir betrachten das in Beispiel 5.10 erläuterte Modell:  $n \geq 3$  stoch. unabhängige normalverteilte Zufallsvariablen  $Y_1, Y_2, \dots, Y_n$ , wobei  $Y_i$  für jedes  $i = 1, 2, \dots, n$  normal- $(y(t_i), \sigma)$ -verteilt ist mit

$$y(t_i) = a t_i + c .$$

Die Werte  $t_1, t_2, \dots, t_n \in \mathbb{R}$  sind bekannt, während  $a, c \in \mathbb{R}$  und  $\sigma > 0$  unbekannt sind und den drei-dimensionalen Parameter  $\vartheta = (a, c, \sigma)$  bilden. Die vorliegenden Beobachtungswerte der Zufallsvariablen  $Y_1, Y_2, \dots, Y_n$  seien mit  $y_1, y_2, \dots, y_n$  bezeichnet. Gemäß Definition 6.4 und Beispiel 5.10 erhalten wir die Likelihood-Funktion

$$L(a, c, \sigma) = \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^n \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - a t_i - c)^2 \right] ,$$

und die Log-Likelihood-Funktion

$$\ell(a, c, \sigma) = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - a t_i - c)^2 . \quad (6.1)$$

Hinsichtlich der Maximierung von  $\ell(a, c, \sigma)$  sehen wir sofort: Die Werte für  $a$  und  $c$  müssen so gewählt werden, dass

$$\sum_{i=1}^n (y_i - a t_i - c)^2 \quad (6.2)$$

minimal wird. Mit den Abkürzungen

$$\begin{aligned} \mathbf{t} &= (t_1, \dots, t_n) , \quad \mathbf{y} = (y_1, \dots, y_n) , \\ \bar{t} &= \frac{1}{n} \sum_{i=1}^n t_i , \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i , \\ s_{\mathbf{t}}^2 &= \frac{1}{n} \sum_{i=1}^n (t_i - \bar{t})^2 , \quad s_{\mathbf{t}, \mathbf{y}} = \frac{1}{n} \sum_{i=1}^n (t_i - \bar{t}) (y_i - \bar{y}) \end{aligned}$$

erhalten wir (als Minimallösung) die ML-Schätzungen

$$\hat{a}(\mathbf{y}) = \frac{s_{\mathbf{t}, \mathbf{y}}}{s_{\mathbf{t}}^2} \quad \text{und} \quad \hat{c}(\mathbf{y}) = \bar{y} - \hat{a}(\mathbf{y}) \bar{t}. \quad (6.3)$$

Dies sieht man wie folgt: (6.2) ist eine quadratische Funktion der zwei reellen Variablen  $a$  und  $c$ :

$$\begin{aligned} \sum_{i=1}^n (y_i - at_i - c)^2 &= \sum_{i=1}^n (y_i^2 - 2at_i y_i - 2y_i c + 2t_i ac + a^2 t_i^2 + c^2) \\ &= \sum_{i=1}^n y_i^2 - 2a \sum_{i=1}^n t_i y_i - 2c \sum_{i=1}^n y_i + 2ac \sum_{i=1}^n t_i + a^2 \sum_{i=1}^n t_i^2 + nc^2. \end{aligned}$$

Die Lösung des Minimierungsproblems erhalten wir durch Bildung der partiellen Ableitungen nach  $a$  und nach  $c$  und “Null-Setzen” derselben. Das liefert (nach kleinen Umformungen) ein lineares Gleichungssystem:

$$\begin{aligned} \left( \sum_{i=1}^n t_i^2 \right) a + \left( \sum_{i=1}^n t_i \right) c &= \sum_{i=1}^n t_i y_i, \\ \left( \sum_{i=1}^n t_i \right) a + n c &= \sum_{i=1}^n y_i. \end{aligned}$$

Die Lösung dieses Gleichungssystems ist:

$$\hat{a}(\mathbf{y}) = \frac{n \sum_{i=1}^n t_i y_i - \left( \sum_{i=1}^n t_i \right) \left( \sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n t_i^2 - \left( \sum_{i=1}^n t_i \right)^2}, \quad (6.4)$$

$$\hat{c}(\mathbf{y}) = \frac{\left( \sum_{i=1}^n t_i^2 \right) \left( \sum_{i=1}^n y_i \right) - \left( \sum_{i=1}^n t_i \right) \left( \sum_{i=1}^n t_i y_i \right)}{n \left( \sum_{i=1}^n t_i^2 \right) - \left( \sum_{i=1}^n t_i \right)^2}. \quad (6.5)$$

Um Formel (6.3) aus den Formeln (6.4) und (6.5) zu gewinnen, verwende man die folgenden beiden Identitäten:

$$s_{\mathbf{t}}^2 = \frac{1}{n} \sum_{i=1}^n t_i^2 - (\bar{t})^2, \quad s_{\mathbf{t}, \mathbf{y}} = \frac{1}{n} \sum_{i=1}^n t_i y_i - \bar{t} \bar{y}.$$

Desweiteren gilt offensichtlich

$$\sum_{i=1}^n t_i = n \bar{t}, \quad \sum_{i=1}^n y_i = n \bar{y}.$$

Damit schreibt sich  $\hat{a}(\mathbf{y})$  aus Formel (6.4) als

$$\hat{a}(\mathbf{y}) = \frac{n^2 \left[ \frac{1}{n} \sum_{i=1}^n t_i y_i - \bar{t} \bar{y} \right]}{n^2 \left[ \frac{1}{n} \sum_{i=1}^n t_i^2 - (\bar{t})^2 \right]} = \frac{s_{\mathbf{t}, \mathbf{y}}}{s_{\mathbf{t}}^2},$$

und  $\hat{c}(\mathbf{y})$  aus Formel (6.5) schreibt sich als

$$\begin{aligned} \hat{c}(\mathbf{y}) &= \frac{n^2 \left[ \left( \frac{1}{n} \sum_{i=1}^n t_i^2 \right) \bar{y} - \bar{t} \left( \frac{1}{n} \sum_{i=1}^n t_i y_i \right) \right]}{n^2 s_{\mathbf{t}}^2} = \frac{\left( s_{\mathbf{t}}^2 + (\bar{t})^2 \right) \bar{y} - \bar{t} (s_{\mathbf{t}, \mathbf{y}} + \bar{t} \bar{y})}{s_{\mathbf{t}}^2} \\ &= \frac{s_{\mathbf{t}}^2 \bar{y} - \bar{t} s_{\mathbf{t}, \mathbf{y}}}{s_{\mathbf{t}}^2} = \bar{y} - \frac{s_{\mathbf{t}, \mathbf{y}}}{s_{\mathbf{t}}^2} \bar{t} = \bar{y} - \hat{a}(\mathbf{y}) \bar{t}. \end{aligned}$$

Es bleibt noch die ML-Schätzung für  $\sigma$  zu berechnen, d. h. wir haben (nach Einsetzen von  $\hat{a}(\mathbf{y})$  und  $\hat{c}(\mathbf{y})$ ) die Maximalstelle  $\hat{\sigma}$  der Log-Likelihood-Funktion

$$\ell(\hat{a}(\mathbf{y}), \hat{c}(\mathbf{y}), \sigma) \quad \text{für alle } \sigma > 0$$

zu bestimmen. Als Abkürzung bezeichne

$$RSS(\mathbf{y}) = \sum_{i=1}^n (y_i - \hat{a}(\mathbf{y}) t_i - \hat{c}(\mathbf{y}))^2,$$

wobei  $RSS$  für *Residual Sum of Squares* steht. Damit schreibt sich die zu maximierende Funktion als (s. Formel (6.1) oben):

$$\ell(\hat{a}(\mathbf{y}), \hat{c}(\mathbf{y}), \sigma) = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) - \frac{1}{2\sigma^2} RSS(\mathbf{y}).$$

Ähnlich wie in Beispiel 6.9 erhalten wir als ML-Schätzung

$$\hat{\sigma}(\mathbf{y}) = \sqrt{\frac{1}{n} RSS(\mathbf{y})}.$$

**Bemerkung 6.11 (Methode der Kleinsten Quadrate)**

Die ML-Methode zur Schätzung der Regressionsparameter  $a$  und  $c$  in Beispiel 6.10 führte zur Minimierung der Summe der Abweichungsquadrate (Formel (6.2)), d. h. die ML-Schätzungen  $\hat{a}(\mathbf{y})$  und  $\hat{c}(\mathbf{y})$  sind die Lösungen des Minimierungsproblems

$$\sum_{i=1}^n (y_i - (a t_i + c))^2 \longrightarrow \min_{a, c \in \mathbb{R}} !$$

Es wird also diejenige Gerade  $y(t) = a t + c$  den Datenpunkten

$$(t_1, y_1), (t_2, y_2), \dots, (t_n, y_n)$$

*angepasst*, welche im Sinne der Summe der Abweichungsquadrate,

$$\sum_{i=1}^n (y_i - y(t_i))^2,$$

die geringste Abweichung von den Datenpunkten liefert (s. auch Abbildung 6.2).

Die ML-Methode zur Schätzung der Regressionsparameter  $a$  und  $c$  stimmt im Normalverteilungsmodell von Beispiel 6.10 also mit der *Methode der Kleinsten Quadrate* überein. Die schließlich verbleibende minimale Abweichungsquadratsumme (*Residual Sum of Squares*),

$$RSS(\mathbf{y}) = \sum_{i=1}^n (y_i - [\hat{a}(\mathbf{y}) t_i + \hat{c}(\mathbf{y})])^2,$$

wird bei der ML-Methode zur Schätzung der Standardabweichung  $\sigma$  verwendet:

$$\hat{\sigma}(\mathbf{y}) = \sqrt{\frac{1}{n} RSS(\mathbf{y})}.$$

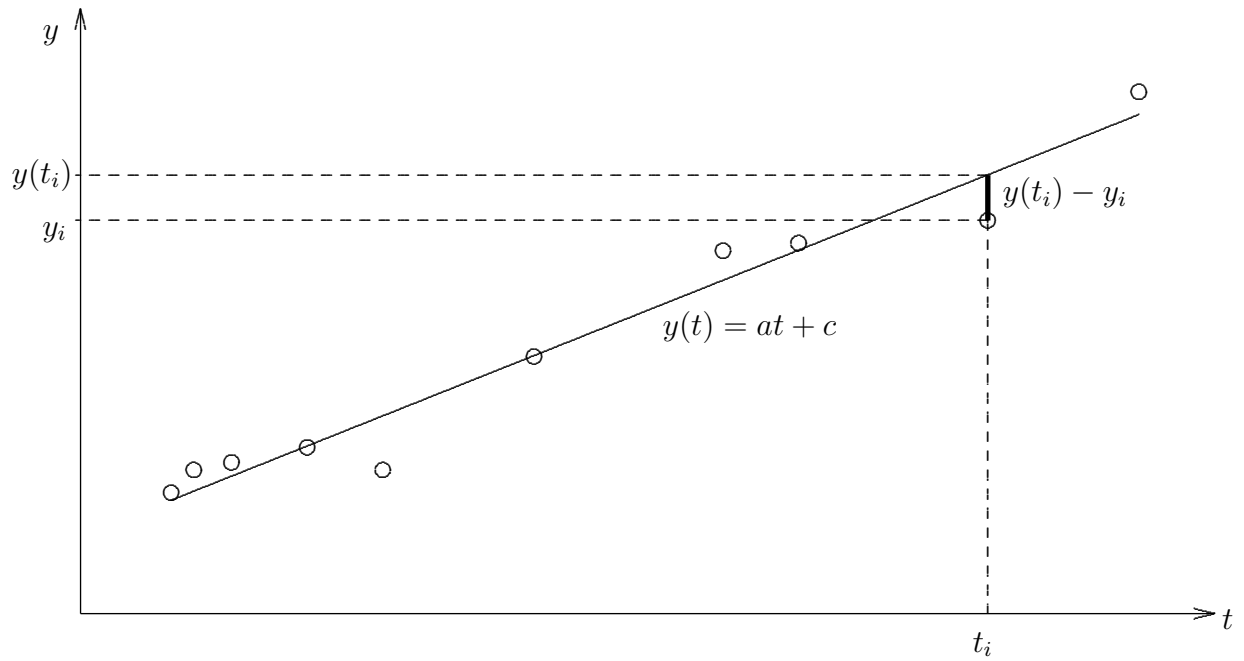


Abb. 6.2: Methode der kleinsten Quadrate: Achsenabschnitt  $c$  und Steigung  $a$  der Geraden werden so bestimmt, dass die Summe der quadrierten vertikalen Abstände der Punkte von der Geraden minimal wird

# Anhang

**Tafel A.1: Verteilungsfunktion  $\Phi$  einer standard-normalverteilten Zufallsvariablen**

$z$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Ablesebeispiel:  $\Phi(0.41) = 0.6591$ .

**Tafel A.2:  $p$ -Quantile  $u_p$  einer standard-normalverteilten Zufallsvariablen**

$p$	0.9	0.95	0.975	0.99	0.995	0.999
$u_p$	1.282	1.645	1.960	2.326	2.576	3.090

**Tafel A.3:  $p$ -Quantile einer  $t_m$ -verteilten Zufallsvariablen**

$m \setminus p$	0.9	0.95	0.975	0.99	0.995	0.999
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
60	1.296	1.671	2.000	2.390	2.660	3.232
70	1.294	1.667	1.994	2.381	2.648	3.211
80	1.292	1.664	1.990	2.374	2.639	3.195
90	1.291	1.662	1.987	2.368	2.632	3.183
100	1.290	1.660	1.984	2.364	2.626	3.174
120	1.289	1.658	1.980	2.358	2.617	3.160
$\infty$	1.282	1.645	1.960	2.327	2.576	3.091



Tafel A.4:  $p$ -Quantile einer  $\chi_m^2$ -verteilten Zufallsvariablen

$m \backslash p$	0.005	0.01	0.025	0.05	0.1	0.90	0.95	0.975	0.99	0.995
1	0.0000	0.0002	0.0010	0.0039	0.02	2.71	3.84	5.02	6.63	7.88
2	0.010	0.020	0.051	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.072	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.65
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.43	104.21
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	107.57	113.15	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	118.50	124.34	129.56	135.81	140.17